LINEAR ALGEBRA

Books by the Author :

Abstract Algebra, Vols. I & II

Linear Algebra

Projective Geometry (In Press)

Topology (In Press)

LINEAR ALGEBRA

BY

B. C. CHATTERJEE

Lecturer in Pure Mathematics, Calcutta University

DAS GUPTA & Co. (Private) Ltd.
54/3, COLLEGE STREET, CALCUTTA-12

Copyright by the Author

All Rights Reserved

This book or any part thereof must not be reproduced in any form without the written permission of the Author.

Price : Rs 20/00

Printed by : Sree Narendra Nath H
Debi Press
33/D, Madan Mitra Lane, Calcutta
Bound by : Sree Guru Binding W
18, Halwell Lane, Calcutta-9.

To

Dr. A. C. Choudhury, M.SC., D. PHIL.

to whose encouragement and inspiration, I owe a lot.

PREFACE

This book has grown out of lectures delivered by the author in the Post Graduate Classes in Pure Mathematics at the University of Calcutta for more than a decade. It is intended to be a text book meant for the Honours and Post Graduate students of the Indian Universities. Topics have accordingly been selected and developed. It may also be profitably used by those who are interested in the application of this subject. Applications of the results of Linear Algebra in Projective, Affine and Metric Geometries have been indicated Almost every topic has been illustrated by worked out examples. Exercises, in a good number, have also been set at the end of each article.

Much of the materials of this book was included in Chapters IV and VII of the first edition of the author's book on Abstract Algebra. These topics have been left out in the revised second edition of that book. Additional materials for this book have been collected from the standard books on the subject. Special mention may be made of Mal'cev's Foundations of Linear Algebra and Gel'fand's Lectures on Linear Algebra. A list of the books, which have been used, is given in the Bibliography at the end. To the authors of all those books, I express my indebtedness.

In spite of best efforts, some annoying misprints have crept in. Any corrections or suggestions for improvement of the book will most thankfully be received by the author,

Calcutta,
1st July, 1967.

B. C. Chatterjee

CONTENTS

I. PRE-REQUISTES 1

1. Sets and subsets, 1
2. Equivalence relation and partition of a set, 3
3. Mappings, 5
4. Systems of single composition, 7
5. Systems of double composition, 11
6. Homomorphisms, 17
7. Theory of divisibility, 18
8. Polynomial rings, 20
9. Algebraic extensions of a field, 24
10. Remarks, 25

II. LINEAR SPACES 26

1. Linear space and subspaces, 26
2. Finite dimensional linear spaces, 33
3. Normed linear spaces, 42

3.1. Euclidean linear spaces, 53

Exercises, 32,40,51,54

III. MATRICES 55

1. Operations on matrices, 55
2. Equivalence of matrices, 62
3. Row and column spaces of a matrix, 66
4. Square matrices, 72

4.1. Complex matrices, 84

Exercises, 61,66,72,81,86

IV. DETERMINANTS 88

1. Determinants, 88
2. Cofactors and minors, 97
3. Minors and rank, 106

Exercises, 95,103,108

V. SOLUTION OF SYSTEM OF LINEAR EQUATIONS...109

1. Homogeneous system, 109
2. Non-homogeneous system, 112
   Intersection of planes in Euclidean 3-space, 116
4. Method of determinants, 118
   Exercises, 112,115,118,120

VI. EQUIVALENCE OF MATRICES 121

1. Equivalence of matrices over a field, 121
2. Equivalence of bilinear forms, 122
3. Equivalence of matrices over a Euclidean domain, 123
   Exercises, 122,123,131

VII. CONGRUENCE OF MATRICES 132

1. Congruence of symmetric matrices, 132
2. Quadratic forms, 136
2.1. Definite, semi-definite and indefinite forms, 142
2.2. Application (Projective classification of quadrics), 144
2.3. Congruence of symmetric bilinear forms, 147
3. Congruence of skew-symmetric matrices, 148
4. Hermitian congruence, 150
   Exercises, 136,142,143,146,147,149,152

VIII. EIGEN VALUES AND EIGEN VECTORS ... 153

1. Matric polynomials, 153
2. Characteristic polynomial, 156
3. Minimal polynomial, 158
5. Eigen values of a matrix, 161
5. Eigen vectors of a matrix, 165
5.1. Application (Fixed points of a collineation of a projective space), 174
   Exercises, 155,157,161,164,174,182

IX. SIMILARITY OF MATRICES 183

1. Similarity of matrices, 183
2. Similarity to a diagonal matrix, 186
3. Block-diagonal matrices, 191

4. Normal forms, 196
5. Unitary similarity, 208
6. Orthogonal similarity, 215
6.1. Application (Classification of quadrics in the Euclidean space), 220
7. Simultaneous reduction of a pair of real symmetric matrices, 228
Exercises, 185, 191, 196, 207, 214, 219, 227, 230

X. LINEAR OPERATORS 231

1. Linear operators on a linear space, 231
2. Rank and nullity of a linear operator, 237
3. Invariant subspaces, 240
4. Adjoint of a linear operator, 249
5. Hermitian and skew-Hermitian operators, 251
6. Unitary and orthogonal operators, 258
7. Spectral decomposition, 266
Exercises, 236, 240, 248, 251, 257, 266, 273

XI MULTILINEAR FUNCTIONS 275

1. Linear functions, 275
2. Dual space, 279
3. Multilinear functions, 286
4. Tensors, 289
4.1. Symmetric and skew-symmetric tensors, 296
Exercises, 278, 285, 289, 295, 299.

BIBLIOGRAPHY

INDEX

CHAPTER I

PRE-REQUISITES

1. Sets and subsets.

By a *class* or *a set* or an *aggregate* is meant a particular collection of objects. The objects are called the *elements* of the set. If an element a belongs to a set A, it is expressed symbolically by $a \varepsilon A$; when an element a does not belong to a set A, one expresses this by $a \notin A$.

If a set B does not contain any element which is not an element of V, then B is said to be a *subset* of A, or that B is *included in* or *contained in* A, or equivalently A is said to *contain* B or *include* B; this is expressed symbolically by $B \subset A$, or $A \supset B$. In particular, $A \subset A$, and A is said to be an *improper subset* of itself. Any other subset of A is called a *proper subset* of A. The *empty set* or the *null set* Φ is conceived as the set which contains no element at all; it is a subset of every set. Two sets A and B are identical if, and only if, $A \subset B$ and $B \subset A$; this is expressed symbolically by $A = B$.

Intersection of sets Let $[P]$ be a given collection of sets. Then the set of all those elements, which are common to all the sets of the collection $[P]$, is called the *intersection* or the *meet* of the collection $[P]$, and is denoted by $\cap[P]$. Intersection of two sets P and Q is denoted by $P \cap Q$.

Two sets P, Q having no element in common, i.e., for which $P \cap Q = \Phi$, are said to be *disjoint*.

Union of sets. By the *union* or *join* of two sets P and Q is meant the set which consists of all those elements, which belong either to P or to Q (may as well belong to both). The union of P and Q is denoted by $P \cup Q$. For any given collection of sets $[P]$, their union $\cup[P]$ is determined by the property that an element belongs to the union if, and only if, it belongs to at least one of the sets P of the given collection $[P]$.

The following properties may now be readily verified, for arbitary sets A, B, C :

(*a*) If $A \subset B$ and $B \subset C$, then $A \subset C$. (*Transitive property*)

(*b*) Any one of the following three relations implies the other two : $A \subset B$, $A \cap B = A$, and $A \cup B = B$. (*Consistency property*)

(*c*) $A \cap A = A$, $A \cup A = A$. (*Idempotent properties*)

(*d*) $A \cap B = B \cap A$, $A \cup B = B \cup A$. (*Commutative properties*)

(*e*) $(A \cap B) \cap C = A \cap (B \cap C)$, $(A \cup B) \cup C = A \cup (B \cup C)$. (*Associative properties*)

(*f*) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$. (*Distributive properties*)

Complement of a subset. Let P be a subset of a set A. A subset Q of A is said to be the *complement of P in A*, if

$$P \cup Q = A \quad \text{and} \quad P \cap Q = \Phi.$$

The complement of a subset P in a set A is usually denoted by $A - P$; and it consists of all those elements of A which do not belong to P.

The following properties, known as *De Morgan's rules*, are of particular importance :

I.1.1. *For any collection of subsets* $[P]$ *of a set* A,

$$A - \cap[P] = \cup[A - P] \quad \textit{and} \quad A - \cup[P] = \cap[A - P].$$

For two arbitrary sets A and B, the *difference set* $A - B$ consists of all those elements of A which do not belong to B.

Cartesian product of sets. Given two non-null sets X and Y, one can form a new set whose elements are the ordered pairs (x, y), where x and y run over the sets X and Y respectively. The set of all such pairs (x, y) is called the *Cartesian product* or *combinatorial product* of the sets X and Y, and is denoted by $X \times Y$.

Similarly, for n non-null sets $P_1,\ldots,P_n$, their Cartesian product $P_1\times\ldots\times P_n$ consists of all ordered n-tuples $(x_1,\ldots,x_n)$, where the kth co-ordinate x_k runs over the set P_k, for $k=1,\ldots n$. If, in particular, $P_1=\ldots=P_n=P$, then the Cartesian product $P_1\times\ldots\times P_n=P\times\ldots\times P$ is denoted by P^n, and called the nth *Cartesian power* of the set P. P^n consists of all n-tuples with all co-ordinates from P.

The definition of the Cartesian product may as well be extended to the case of any infinite aggregate of non-null sets. For an *enumerable* aggregate of sets $P_1, P_2,\ldots,P_n,\ldots$, the Cartesian product $P_1\times P_2\times\ldots\times P_n\times\ldots$ consists of all infinite sequences $(x_1, x_2,\ldots,x_n,\ldots)$, where the kth co-ordinate x_k runs over the kth set P_k, for $k=1, 2,\ldots,n,\ldots$. In particular, the Cartesian power P^d of a non-null set P, where d is the cardinal number (or potency) of an enumerable set, consists of all infinite sequences with elements from P.

2. Equivalence relation and partition of a set.

By a *binary relation* ρ defined for the elements of a set A, one means that the proposition "x has the relation ρ to y" is either *true* of *false*, for each pair of elements x, y belonging to A. When the proposition is true, one expresses this by the *propositional function* $x\rho y$. Unless otherwise stated, a relation will mean a binary relation.

A relation ρ defined for the elements of a set A is said to be an *equivalence relation*, or an *RST relation*, if the following properties hold :

(i) $a\rho a$, for every a belonging to A. (*Reflexive property*)

(ii) $a\rho b$ implies $b\rho a$, where $a, b \in A$. (*Symmetric property*)

(iii) $a\rho b$ and $b\rho c$ imply $a\rho c$, where $a, b, c \in A$. (*Transitive property*)

The importance of an equivalence relation is revealed by the following theorem :

I.2.1. *An equivalence relation ρ, defined for the elements of a given set A, determines a partition of the set A into disjoint* *classes, such that two elements a, a' of A belong to the same class if, and only if, $a\rho a'$ holds.*

Conversely, given any partition of the set A, let ρ donote the relation, such that $a\rho a'$ holds when a' belongs to the same class as a, then ρ is an equivalence relation

Proof: Let ρ be an equivalence relation defined for the elements of a set A. All those elements a', for which $a\rho a'$ holds, form a subset of A; this is denoted by class (a). If a' and a'' belong to (a), then $a\rho a'$ and $a\rho a''$ hold. $a\rho a'$ implies $a'\rho a$, by the symmetric property. Then $a'\rho a$ and $a\rho a''$ imply $a'\rho a''$ by the transitive property. If (a) be not identical with A, then there exists an element b of A which does not belong to (a), i.e., $a\rho b$ does not hold. The class (b) is then defined to be the subset consisting of all those elements b' of A for which $b\rho b'$ holds. Then b' does not belong to (a), i.e., $a\rho b'$ does not hold; otherwise, $a\rho b'$ and $b\rho b'$ would imply $a\rho b$ (by the symmetric and transitive properties of ρ), which contradicts the condition that $a\rho b$ does not hold. Hence the classes (a) and (b) are disjoint; and $a'\rho b'$ does not hold for any $a' \varepsilon (a)$ and $b' \varepsilon (b)$. If the elements of A be not all contained in the two classes (a) and (b), the same process is repeated to obtain new classes; and the required partition of the set A is thus obtained.

Conversely, when a partition of a set A into disjoint classes is given, a relation ρ is defined for the elements of A, such that $a\rho a'$ means that a' belongs to the same class as a. The three properties of reflexivity, symmetry and transitivity can then be readily verified for ρ. Hence ρ is an equivalence relation.

The disjoint classes into which a set A is partitioned by an equivalence relation ρ constitute a set, called the *quotient set of A with respect to ρ.*

The set of residue classes $J/(g)$. Consider the set J of all integers, and let $a\rho b$ mean that "$a-b$ is divisible by a fixed positive integer, say g" [this is also expressed symbolically by $a \equiv b \pmod{g}$]. Then ρ is an equivalence relation. $a-a=0$ is divisible by g, hence $a\rho a$ holds for every $a \varepsilon J$ – thus ρ is reflexive. If $a-b$ be divisible by g, then $b-a$ is also divisible by g; hence $a\rho b$ implies $b\rho a$ – thus ρ is symmetric. Finally, if $a-b$ and $b-c$ be both divisible by g, then so also is their sum $a-c$; hence $a\rho b$ and $b\rho c$ imply $a\rho c$ – thus ρ is transitive.

When any integer a, positive, negative or zero, is divided by g, there are g possible remainders, viz. $0, 1, \ldots\ldots, g-1$. If r be the remainder for a, then $a-r$ is divisible by g, so that $a \varepsilon (r)$. Hence every integer belongs to one and only one of the g classes $(0), (1), \ldots, (g-1)$. Two integers a, a' belong to the same class if and only if $a-a'$ is divisible by g, i e., if and only if they leave the same remainder on division by g. Thus the quotient set of J with respect to this equivalence relation consists of the g distinct classes, $(0), (1), \ldots, (g-1)$; the quotient set is denoted by $J/(g)$, and called the *set of residue classes modulo g*.

3. Mappings.

A *correspondence* between the elements of two given sets X and Y, such that for each element $x \varepsilon X$ there correspond some definite elements in Y, is called a *mapping of X into Y* (or *on Y*), or a *function on X to Y*. If the mapping be denoted by β, then the elements of Y which correspond to the element $x \varepsilon X$ will be denoted by $x\beta$, and called the *images of x under β*, or the *values of β at x*; these form a subset of Y, denoted $X\beta$. The set X is called the *range of β*, and $X\beta$ is called the *domain of values of β on X*. If $X\beta$ coincides with Y, β is said to be a mapping of *X onto Y*.

A mapping β is said to *single-valued* if, for every element $x \varepsilon X$, the set $x\beta$ consists of a single element only; otherwise, the mapping is said to be *multiple-valued.* A single-valued mapping β of a set X into a set Y is said to be

(i) *injective* if $x \neq x'$ in X implies $x\beta \neq x'\beta$ in Y;
(ii) *surjective*, if $X\beta = Y$, i.e., β is an *onto* mapping;
(iii) *bijective*, if it is both injective and surjective.

Injective, surjective and bijective mappings are called *injections*, *surjections* and *bijections* respectively. An injection of X

on Y is also known as a *biuniform mapping* of X into Y, whereas a bijection is called a *biuniform mapping* of X onto Y.

Let α be a single-valued mapping of a set X into a set Y, and let M be a subset of Y. The totality of all elements x in X, for which $x\alpha \varepsilon M$, form a subset of X, called the *inverse image of M under* α, and is denoted by $M\alpha^{-1}$. Since for an element $y \varepsilon Y$, there may not exist any *element* $y\alpha^{-1}$ belonging to X, α^{-1} cannot be considered as a mapping of Y into X, unless α is surjective.

Let α be a surjection of X onto Y. The mapping μ of Y onto X, by which an element $y \varepsilon Y$ is mapped on those elements $x \varepsilon X$ for which $x\alpha = y$, is said to be the inverse of α, and is denoted by α^{-1}. The mapping α^{-1} may not however be single-valued. When α^{-1} is also single-valued, the mappings α and α^{-1} are both bijective. In this case, every $x \varepsilon X$ has a unique image $y \varepsilon Y$, and every $y \varepsilon Y$ is the image of a unique element $x \varepsilon X$. Thus α establishes a *biunique* or (1, 1)*-correspondence* between the elements of X and Y.

It now follows readily from the above definitions that :

I.3.1. *A single-valued mapping α of a set X onto a set Y induces an equivalence relation ρ for elements of X, such that $x\rho x'$ holds if, and only if, $x\alpha = x'\alpha$; and one obtains a biuniform mapping of the quotient set of X, with respect to ρ, onto Y.*

Transformations. Let X be a given set. A single-valued mapping β of X into X is called an *operator*, or an *endomapping*, or more commonly a *transformation* of the set X. A biuniform mapping of X onto X is called a *non-singular transformation* or a *permutation* of X, the term permutation being generally used for finite sets only. A transformation, which is not non-singular, is said to be *singular*.

Compositions. Let A be a given set. A single-valued mapping of $A \times A$ into A is called a *composition* or a *binary operation* in A ; and A is said to be *closed with respect to the composition*. Thus, by a composition, for every ordered pair of elements u, v of A, a unique element, say w, is determined in A ; w is called the *composed of u and v*.

Let a composition be defined in a set A, and let B be a subset of A. Though the term "composition" is used for elements of B, and the composition is supposed to be defined in B also, the composed of a pair of elements belonging B may not lie in B. In that case, B is *not closed* with respect to the composition.

The possible compositions in a set naturally depends on the nature of the particular set ; and it is generally possible to define many different and independent compositions in the same set.

4. Systems of single composition.

A set R, in which a composition has been defined, is called a *system of single composition.* Let the composition be called *multiplication*, and let multiplicative terms and notations be used ; then

(i) R is said to be *closed with respect to the composition*, if for every pair of elements $a, b \varepsilon R$, their composed $a.b$ or ab (called the *product* of a, b) belongs to R ; (*Closure property*)

(ii) the elements of R are said to be *associative*, with respect to the composition, if $(ab)c=a(bc)$ for every triplet $a, b, c \varepsilon R$; (*Associative property*)

(iii) the elements of R are said to be *commutative*, with respect to the composition, if $ab=ba$, for every pair $a, b \varepsilon R$; (*Commutative property*)

(*iv*) the *laws of inverse existence* are said to be satisfied in R, if for every pair $a, b \varepsilon R$, there always exist unique elements $x, y \varepsilon R$, such that $ax=b$ and $ya=b$. (*Property of inverse existence*)

The above four properties are independent of each other ; that is, there exist different systems in which some of these properties are satisfied and others do not hold good. But for every system, which is of some practical importance, the closure property (i) is satisfied ; in fact, this is taken to be implied in the definition of the composition.

A system R is said to be a *groupoid*, if the closure property holds in it. A groupoid, in which the associative property (ii) also holds, is called a *semigroup.* While a groupoid, in which the property of inverse existence (iv) is satisfied, is called a *quasigroup.*

A groupoid, in which both the associative property (ii) and the property of inverse existence (iv) are satisfied, is called a *group.* A group, in which the commutative property (iii) holds, is called a *commutative group* or an *abelian group.*

Possible relations between these different systems of single composition are treated elaborately in books on abstract algebra.

Groups and subgroups. Let G be a group, the composition being taken as multiplication. Then for any element $a \varepsilon G$, there exists, by the property of inverse existence, an element $e \varepsilon G$, such that $ae=a$. If b be an arbitary element of G, then by the property of inverse existence there exists a unique element $c \varepsilon G$, such that $ca=b$. Then

$$\begin{aligned} be &= (ca)e = c(ae) && \text{(by the associative property)} \\ &= ca = b. \end{aligned}$$

Thus. for the element e, $ae=a$ *holds for all elements* $a \varepsilon G$. Similarly, there exists an element $f \varepsilon G$, such that $fa=a$, for all elements $a \varepsilon G$. Then, $fe=e$, by the property of f; and $fe=f$, by the property of e, Hence $e=f$, and one obtains ;

I.4.1. *In a group G, there exists a unique element e such that*

$$ae=a=ea, \quad \textit{for all elements } a \varepsilon G.$$

The distinguished element e (of the above theorem) is called the *singular element* of the group G. In multiplicative terminology, the singular element is called the *identity* element, and is denoted by 1.

Again, for a given element $a \varepsilon G$, there exists, by the property of inverse existence, an element a', such that $aa'=1$.

It can be shown that $a'a=1$ also holds, and that a' is unique. It is called the *inverse of* a, and is denoted by a^{-1}. Thus a^{-1} is characterised by the property $a^{-1}a=1=aa^{-1}$. The following *cancellation properties* can now be readily proved :

(i) $au=av$ *implies* $u=v$, and (ii) $pa=qa$ *implies* $p=q$, where $a, u, v, p, q \varepsilon G$. In fact, $u=a^{-1}(au)=a^{-1}(av)=v$, and $p=(pa)a^{-1}=(qa)a^{-1}=q$.

A subset H of a given group G is said to be a *subgroup* of G, if H forms by itself a group with respect to the same composition as in G. It is evident that the identity element of G belongs to H, and is the identity element of H. The following necessary and sufficient conditions for a subgroup can readily be established.

I.4.2. *A subset H of a multiplicative group G is a subgroup of G if, and only if, (i) the product of any two elements of H always belongs to H, and (ii) the inverse of every element of H belongs to H.*

Proof : The necessity of the conditions is obvious. As to the sufficiency of the conditions, the given condition (i) ensures the closure property, the associative property is hereditary, and the solutions of $ax=b$ and $ya=b$ are given by $x=a^{-1}b$ and $y=ba^{-1}$ (which are unique by the cancellation properties).

An alternative condition in the following form can also be readily obtained :

I.4.3. *A subset H of a multiplicative group G is a subgroup of G if, and only if $ab^{-1} \varepsilon H$, for every pair of elements $a, b \varepsilon H$.*

It can be shown that *for a finite group, the condition (i) of* 1.4.2 *alone is sufficient for a subset H to be a subgroup of G.*

It now follows immediately from I.4.2 or I.4.3 that :

I.4.4 *The intersection of any aggregate of subgroups $[P]$ of a group G is a subgroup of G.*

Proof: Let M be the intersection of the collection $[P]$. Let $a, b \varepsilon M$, then a, b belong to *each* P ; and since P is a subgroup, $ab^{-1} \varepsilon P$. Thus ab^{-1} belongs to each P in $[P]$, so that $ab^{-1} \varepsilon M$. Hence, by I.4.3 M is a subgroup of G.

Let a be a given element of a group G, then the powers of a are defined by

$$a^0=1,\ a^1=a,\ a^{n+1}=a^n.a,\ \text{and}\ a^{-n}=(a^n)^{-1},$$

where n is any positive integer ; n is said to be the *index* of a^n. Thus powers of a are defined for all integral values (positive, negative and zero) of the index. The laws of indices, $a^r.a^s={}^{r+s}$ and $(a^r)^t=a^{rt}$, hold for all integral values of r, s, t. It now follows readily, by I.4.2, that all powers of a form a subgroup of G ; it is called the *cyclic subgroup*, generated by the element a. The cyclic subgroup, generated by an element a, is the intersection af all subgroups of G containing the element a.

If, for an element a, there exists a positive integer m, such that $a^m=1$, then the smallest positive integer m, for which $a^m=1$, is called the *order of the element a*. If, on the other hand, there does not exist any positive integer m satisfying $a^m=1$, then the element a is said to be of *order* 0 or *infinity*. It can be easily shown that for an element a of order m, the cyclic subgroup generated by a contains axactly m distinct elements, whereas for an elements a of order infinity, the number of distinct elements in the cyclic subgroup generated by a is infinitely many. The number of elements in a group (or a subgroup) is said to be its *order.* It then follows that the order of the cyclic subgroup generated by an element a is the same as the order of a.

Let H be a given subgroup of a multiplicative group G. Let a relation ρ be defined for the elements of G, such that $a\rho b$ holds for two elements a, $b \varepsilon G$, if $a^{-1} b \varepsilon H$. Then ρ is an equivalence relation Hence G is partitioned into disjoint classes ; these classes are called the *left cosets of H in G.* Similarly, the *right cosets of H in G are* determined by an equivalence relation σ, where $a\sigma b$ holds for two elements a, $b \varepsilon G$, if $ba^{-1} \varepsilon H$. An element b belongs to the left coset, containing an element a, if and only if, $a\rho b$ holds, i.e., $a^{-1}b \varepsilon H$ i.e., $b \varepsilon aH$ (where aH consists of all products ah, as h runs over H) ; thus aH denotes the left coset which contains the element a. Likewise the right coset containing an element a is given by Ha. In general, aH is different from Ha.

A subgroup N is said to be a *normal subgroup* of a group G, if every left coset of N in G is also a right coset (and conversely) i.e., if $aN=Na$, for every element $a \varepsilon G$.

By defining the product of (left) cosets by $aH.\ bH=(ab)\ H$, it can be shown that the (left) cosets of a subgroup H in a group G form a group if, and only if, H is a normal subgroup of G ; and then the group formed by cosets is called the *factor group* G/H.

G and the identity subgroup I are always normal subgroups of any group G ; these are distinguished as the two *trivial normal subgroups* of G. A group G, which has no normal subgroups, besides the two trivial ones, is called a *simple group.*

Again $h\alpha=ah$ is a biuniform mapping of the subgroup H onto the left coset aH. Thus there exists a one-to-one correspondence between the elements of H and those of any left coset of H. If G be a group of finite order n, and if the order of H be r, then every left coset of H also contains r distinct elements. Since the left cosets are disjoint, it follow that $n=rs$, where s is the number of left cosets of H in G (s is called the *index of H in G*). Thus one obtains :

I.4.5 Lagrange's theorem. *For a finite group G, the order of any subgroup is a divisor of the order of G.*

Since the order of any element is the same as the order of the cyclic subgroup generated by that element, it follows immediately from Lagrange's theorem that :

I.4.6. *The order of any element of a finite group G is a divisor of the order of G.*

5. Systems of double composition.

Let R be an additive commutative group. In additive notation, the singular element of a group is called the *null element*, and is denoted by 0 ; and the inverse of an element a is called its *negative*, and is denoted by $-a$. Thus the null element 0, and the negative $-a$ are characterised by the properties :

$$0+g=g=g+0, \quad \text{for all elements } g \in R,$$

and

$$(-a)+a=0=a+(-a).$$

Let a second composition, called multiplication be defined for elements of R. Then the *distributive properties* are said to be satisfied in R, If

$$a(b+c)=ab+ac, \qquad (a+b)c=ac+bc$$

hold, for any triplet a, b, $c \in R$.

If, now, R be a semigroup with respect to muitiplication (it is already a commutiative group with respect to addition), and if the distributive properties are satisfied in R, then it is said to be a *ring.* If the commutative property of multiplication be also satisfied in R, then it is called a *commutative ring.* If further the laws of inverse existence be satisfied in the restricted form, viz. $ax=b$ and $ya=b$ have unique solutions x, y in R provided that $a \neq 0$, then R is called a *field.* Thus a field R is additively a commutative group, whose non-null elements form a multiplicative commutative group, and $0.a=0$ for every $a \in R$. Hence both the null element 0 and the identity 1 element exist in a field, and are unique. A field, in which the commutative property of multiplication is not assumed, is called a *skew field.*

I.5.1. *In a skew field R, $ab=0$ implies that at least one of a and b must be the null element 0.*

Proof : If $a \neq 0$, then $a^{-1} \in R$, and $b=1.b=(a^{-1}a)b=a^{-1}(ab) =a^{-1}.0=0$.

The above property is not necessarily satisfied in an arbitrary ring. In fact, there exist rings in which the product of two non-null elements may be the null element ; such elements are known as the *divisors of zero.* A commutative ring, which does not contain any divisor of zero, is called an *integral domain.* Whereas every field is an integral domain, an integral domain may not be a field. E.g. the integers form an integral domain, which is not a field. But, it can however be shown that *a finite integral domain is necessarily a field.* Similarly, every field is a skew field, but a skew field may not be a field. E. g. the rational quaternions form a skew field, which is not a field. But, it can be shown that *a finite skew field is necessarily a field.*

Characteristic. If m be the smallest positive integer for which $ma=0$ holds [where $ma=a+a+\dots(m$ terms)], for all elements a belonging to a ring R, then m is said to be the *characteristic of the ring R.* If there does not exist any such positive integer, then R has the *charcteristic infinity* or *zero.* The ring J/(g) (defined in page 15) has the characterstic g, while the ring of integers J has the characteristic 0.

It can be shown that *the characteristic of a field is either* 0 *or a prime number p.* The field $GF(p)$ (see page 16) has characteristic p, whereas the characteristic of the field of rational numbers, the field of real numbers and the field of complex numbers is 0.

The binomial theorem, for positive integral index, holds in any commutative ring. And since for a prime p, the binomial coefficients ${}^pC_r$ are multiples of p, for $p=1, 2,\dots,\dots p-1$, it follows that in a field of characteristic p, $(a \pm b)^p = a^p \pm b^p$.

In a field of characteristic 2, $a+a=0$, so that $a=-a$ holds, for every element a.

Subsystems. Let R be a given system of double composition and S a subset of R, then S is said to be a *subgroup* (with respect to either addition or multiplication) or a *subring* or a *subfield* of R, according as S is by itself a group or a ring or a field (respectively), with respect to the same compositions of addition and multiplication as in R. Opposite to the concept of a subsystem is that of an extension. By an *extension* of a group or a ring or a field K is meant respectively a group or a ring or a field M, which contains K as a subgroup or a subring or a subfield respectively. E.g. the set of even integers forms a subring J_2 of the ring of integers J; and so, J is an extension of J_2.

Similarly, the rational numbers from a subfield **R** of the field of real numbers **E**; **E** is therefore an extension of **R**.

Analogous to I.4.2. and I.4.3, one obtains the following:

I.5.2. *A subset S of a ring R is a subring of R if, and only if, $a-b$, $ab \varepsilon S$ for every pair of elements $a, b \varepsilon R$.*

Proof. The given conditions are obviously necessary. To prove the sufficiency, let the conditions hold in S. Since R is additively a group and $a-b \varepsilon S$, for every pair $a, b \varepsilon S$, it follows by I.4.3 that S is an additive subgroup of R. Commutative property of addition and the associative properties of addition and multiplication are hereditary; hence these hold in S. The closure property of multiplication holds in S, by hypothesis. Hence S forms a ring, by itself. Thus S is subring of R.

I.5.3. *A subset K of a field M is a subfield of M if, and only if, $a-b$, $ab \varepsilon K$, when $a, b \varepsilon K$, and $c^{-1} \varepsilon K$, for every non-null $c \varepsilon K$.*

Proof: The necessity of the given conditions is obvious. Conversely, let the given conditions hold in K Then, by I.5.2, it follows that K is a subring of M. Commutative property of multiplication, being hereditary, holds in K. Also the equations $ax=b$ and $ya=b$ have unique solutions $x=a^{-1}b$ and $y=ba^{-1}$ in K, provided that $a \neq 0$. Consequently, K is a field by itself, and is therfore a subfield of M.

I.5.4. *A subring of a field is an integral domain.*

Proof: Let D be a subring of a field K. The commutative property of multiplication, being hereditary, holds in D. Also, as there are no divisors of zero in K, D does not contain any divisor of zero. Hence D is an integral domain.

The converse of I.5.4 is also true. It can be shown that every integral domain D can be extended to a field. In fact, there exists a field $Q(D)$, containing D as a subring, such that $Q(D)$ consists of all quotients ab^{-1}, where $b \neq 0$ and a, b run over D; $Q(D)$ is called the *quotient field of D*. Since, for any field K containing D as a subring all quotients ab^{-1}, where $b \neq 0$ and a,

$b \in D$, are contained in K, it follows that $Q(D)$ is contained in K. Thus *any field K, containing an integral domain D as a subring, also contains the quotient field $Q(D)$ as a subfield.*

Analogous to I.4.4, one can readily deduce the following from I.5.2 and I.5.3 :

I.5.5. (i) *The intersection of any aggregate of subrings of a ring R is a subring of R.*

(ii) The intersection of any aggregate of subfields of a field K is a subfield of K.

Residue class ring J/(g). Addition and multiplication are defined in the set of residue classes J/(g) (vide page 4) by :

$$(a)+(b)=(a)+(b) \text{ and } (a).\,(b)=(ab).$$

The sum and product, are independent of the choice of representatives of the classes. In fact, if $(a')=(a)$ and $(b')=(b)$, then $a'=a+rg$ and $b'=b+sg$, so that

$$a'+b'=a+b+[r+s]g, \text{ and } a'b'=ab+[as+br+rsg]g\ ;$$

hence $$(a'+b')=(a+b) \text{ and } (a'b')=(ab).$$

The associative and commutative properties of addition and multiplication, and the two distributive properties can readily be verified. In fact, these properties follow from the corresponding properties for integers. (0) and (1) are respectively the null and the identity elements of J/(g), since

$$(0)+(a)=(0+a)=(a) \text{ and } (1).(a)=(1.a)=(a),$$

for every element $(a) \in$ J/(g). The negative of (a) is given by $(-a)$, since $(-a)+(a)=(-a+a)=(0)$.

Thus J/(g) *is a commutative ring with the identity element, for any positive integer g.*

If $g=rs$ be a composite number $(1<r<g,\ 1<s<g)$, then $(r) \neq (0)$ and $(s) \neq (0)$, but $(r)(s)=(rs)=(g)=(0)$. Thus (r), (s) are divisors of zero ; hence J/(g) *cannot be a field* (by I.5.1).

On the other hand, *let g be a prime number.* Let $(a) \neq (0)$, then the integer a is not divisible g. Hence the h.c.f. of a and g must be 1, and there exist integers b and c, such that $ab+gc=1$.

Then $ab-1$ being divisible by g, and it follows that $(1)=(ab)=(a)(b)$; i.e., (b) is the inverse of (a). Thus every non-null element of $J/(g)$ has an inverse in $J/(g)$; consequently, $J/(g)$ is a field.

Combining the results, obtained above, one gets the following:

I.5.6 *The residue classes form a commutative ring* $J/(g)$, *with the identity element, for any positive integer* g. *This ring is a field if, and only if,* g *is a prime.*

For a prime p, $J/(p)$ is a field; it is usually denoted by $GF(p)$. Thus there exists a field with p elements, for any prime number p. It can be shown that, *for a given positive integer* n, *there exists a field with* n *elements if, and only if,* $n=p^m$, *where* p *is a prime and* m *is any positive integer; the field with* p^m *elements is denoted by* $GF(p^m)$, *and called the Galois field with* p^m *elements.*

Ideals and quotient rings. A subset P of a ring R is said to be a *both-sided ideal* or simply an *ideal* of R, if for any $u, v, \varepsilon P$ and any $a \varepsilon R$; $u-v$, au and ua belong to P. An ideal is evidently a subring (by I 5.2), but not conversely. As in I.5.5, the intersection of any aggregate of ideals of R is also an ideal of R.

If R be a commutative ring, containing the identity element, then for elements $g_1,\ldots,g_n \varepsilon R$, the set of all elements $a_1g_2+\ldots+a_ng_n$, as $a_1,\ldots, a_n$ run over R independently, form an ideal of R, containing $g_1,\ldots,g_n$; it is denoted by $(g_1,\ldots,g_n)$, and called the *ideal generated by the elements* $g_1,\ldots, g_n$. This ideal is contained in every ideal of R, which contains the elements $g_1,\ldots,g_n$. In particular, the ideal generated by a single element g is called the *principal ideal* (g).

In general, there may exist ideals of R, which are not finitely generated. If every ideal of a commutative ring R, containing the identity element, be finitely generated, then R is called a *Noetherian ring*. If, further, every ideal of R be a principal ideal, then R is called a *principal ideal ring*.

Let P be an ideal of a ring R. Let a relation ρ be defined for the elements of R, such that $a\rho b$ holds, if $a-b \varepsilon P$. Then ρ is an equivalence relation. R is accordingly partitioned into disjoint classes; these are called the *residue classes in R, modulo P.* Let (a) denote the residue class which contains the element a, and let addition and multiplication of residue classes be defined by

$$(a)+(b)=(a+b) \text{ and } (a).(b)=(ab).$$

Then it follows, as in the case of $J/(g)$, that the residue classes form a ring; it is called the *quotient ring R/P.* And, if R be a commutative ring, then the quotient ring R/P is also so.

6. Homomorphisms.

A single-valued mapping α of a ring R onto a system of double composition S is a called a *homomorphism*, if:

$$(a+b)\alpha=a\alpha+b\alpha \text{ and } (ab)\alpha=a\alpha.b\alpha,$$

for every pair of elements $a, b \varepsilon R$. S is then said to be *homomorphic to R*, or a *homomorph* (or a *homomorphic image*) *of R under α.* It then follows that S is also a ring. And, if R be a commutative ring, then so also is S. A homomorphism α of R onto S is called an *isomorphism*, if α is in particular injective (and therefore also bijective). Then α^{-1} is also an isomorphism of S onto R; R and S are therefore isomorphic to each other. An isomorphism establishes a one-to-one correspondence between the elements of R and S, in such a way that to the sum and product of two elements of R, there correspond the sum and product (respectively) of their images in S. Consequently, every property, that holds in R, holds correspondingly in S also. *Two isomorphic systems are therefore abstractly equal.*

Let α be a homomorphism of a ring R onto S. If 0 be the null element of R, then 0α is the null element of S, since $a\alpha+0\alpha=(a+0)\alpha=a\alpha$ holds for every $a\alpha \varepsilon S$. Let P be the subset of R, consisting of all those elements u, for which $u\alpha=0\alpha$. Then P is a both-sided ideal of R; it is called the *kernel of the homomor-*

phism α. It then follows, as in I.3.1, that there exists a biuniform mapping, say β, of the quotient ring R/P onto S; and the mapping β can readily be shown to be, in particular, an isomorphism. Thus one obtains:

I.6.1. Homomorphism theorem: *If S be homomorphic to a ring R, then S is isomorphic to the quotient ring R/P, where P is the kernel of the given homomorphism.*

On the other hand, if P be any both-sided ideal of a ring R, then $a\alpha=(a)$ is a homomorphism of R onto the quotient ring R/P. Hence *the homomorphs of a ring R are all given by the different quotient rings of R.*

The concepts of homomorphism and isomorphism are defined also for groups, where only one composition is taken into account. For multiplicative groups G and H, the kernel N of a homomorphism α of G onto H is a subset of G, consisting of all those element u for which $u\alpha=1\alpha$. Then N is a normal subgroup of H, and one has:

I.6.2. Homomorphism theorem (for groups): *If H be homomorphic to a group G, then H is isomorphic to the factor group G/N, where N is the kernel of the given homomorphism.*

Also, any factor group of a group G is homomorphic to G; and thus *the homomorphs of a group G are all given by the different factor groups of G.*

7. Theory of divisibility.

Let D be an integral domain, containing the identity element 1. An element a is said to be *divisible* by an element b in D, if there exists an element $c \,\varepsilon\, D$, such that $a=bc$. Then b (as also c) is said to be a *divisor* or a *factor* of a, and this relation is denoted symbolically by $b \mid a$.

An element u is said to be a *unit* in D, if u^{-1} also belongs to D. Then u^{-1} is also a unit in D. Product of two units is a unit. And a unit is a divisor of every element of D.

An element a' is said to be *associated* to an element a (or an *associate* of a) in D, if $a'=au$, where u is a unit. It then follows readily that two elements are associated if, and only if, they are mutually divisible in D.

Irreducible and prime elements. Every element is divisible by the units and all its own associates. A non-null, non-unit element h is said to be *irreducible in* D, if it has no divisors except the units and its own associates. A non-null, non-unit element, which is not irreducible, is said to be *reducible in* D.

A non-null, non-unit element p is said to be *prime in* D, if $p \mid ab$ implies $p \mid a$ or $p \mid b$, where $a, b \in D$. It can be proved easily that *every prime element is irreducible, but an irreducible element may not be prime in an integral domain* D.

Factorisation. A non-null, non unit element a is said to be *factorisable* in D, if a can be expressed as a product of some irreducible elements, $a=h_1 \ldots h_r$, where h_i are irreducible, and r may as well be 1. The product expression $h_1 \ldots h_r$ it then called *a factorisation* of the element a. It should however be noted that an element may not at all be factorisable. If every non-null, non-unit element of D be factorisable, then D is called a *factorisable domain.*

Two factorisations, $a=g_1 \ldots g_k$ and $a=h_1 \ldots h_r$, of an element a are said to be equivalent, if $k=r$, and there exists a one-to-one correspondence between the factors $g_1, \ldots, g_k$ and $h_1, \ldots, h_r$, such that every pair of corresponding factors are associated in D. An element a is said to be *uniquely factorisable*, if any two factorisations of a are always equivalent. A factorisable domain D is said to be a *uniquely factorisable domain* (or simply, a *u.f.d.*) or a *Gaussian domain*, if every non-null, non-unit element of D is uniquely factorisable.

It can be shown that *every Noetherian domain* (i.e., a Noetherian ring, which does not contain any divisor of zero) *is necessarily a factorisable domain*, and that *a factorisable domain is a Gaussian domain if, and only if, every irreducible element is also prime in the domain.*

Highest common factor. An element c is a common factor to two given elements a and b, if $c \mid a$ and $c \mid b$. An element h is called a *highest common factor* (or simply *h.c.f.*) or a *greatest common divisor* (or simply *g.c.d.*), if h be a common factor of a and b, and such that every common factor of a and b is always a factor of h. Two highest common factors of a and b are mutually divisible, and are therefore associated.

The *h.c.f.* of two elements a and b can always be found by factorisation, as in elementary mathematics, when D is a Gaussian domain. The method of finding the h.c.f. of two elements by division algorithm, as in elementary mathematics, can also be applied in an abstract integral domain, by imposing certain restrictions, as follows.

Euclidean domain. An integral domain D, in which the following two conditions hold, is called a *Euclidean domain* :

E(1) : A *norm function* N can be defined on D, such that to every non-null element a of D, there corresponds a positive integer aN, and such that $(ab)N \geqslant aN$ (and also $\geqslant bN$).

E(2) : A *division algorithm* can be defined in D, such that for every pair of elements a, b, with $b \neq 0$, there exist two elements q, r in D, satisfying $a = qb + r$, where either $r = 0$, or $rN < bN$.

The ring of integers is the most common example of a Euclidean domain. As in elementary mathematics, one can find the h.c.f. of any pair of non-null elements a, b, belonging to a Euclidean domain D, by applying division algorithm a finite number of times. One can also show that the h.c.f. h of any two elements $a, b \in D$ is expressible as $h = fa + gb$, where f, g are two suitable elements of D.

8. Polynomial rings.

Let R be a given ring. An expression $a_0 + a_1x + \ldots + a_nx^n$, where $a_0, a_1, \ldots, a_n \in R$, and x is an abstract element (not belonging to R), is called a *polynomial* over R ; and a_0, $a_1, \ldots, a_n$ are called its *coefficients*. The element x is called an *indeterminate* or a *variable*. If $a_n \neq 0$, then n is called the *degree of the polynomial*; if, on the otherhand, $a_n = 0$, then the degree is r, where $r < n$, and such that $a_r \neq 0$ but $a_{r+1} = \ldots a_n = 0$. A polynomial, all whose coefficients are null, is called the *null polynomial*; and its degree is defined to be -1.

The set of all polynomials over R is denoted by $R[x]$. Let

$$f(x) = a_0 + a_1 x + .. + a_n x^n \text{ and } g(x) = b_0 + b_1 x + \ldots + b_m x^m$$

be any two polynomials belonging to $R[x]$, and let $m \leqslant n$. The sum and product of the polynomials $f(x)$ and $g(x)$ are defined by

$$f(x) + g(x) = s(x) = c_0 + c_1 x + \ldots + c_n x^n,$$

and
$$f(x).g(x) = p(x) = d_0 + d_1 x + \ldots + d_{n+m} x^{n+m},$$

where
$$c_k = a_k + b_k, \text{ and } d_k = a_0 b_k + a_1 b_{k-1} + \ldots + a_k b_0,$$

for $k = 0, 1, \ldots$, taking $b_{m+1} = \ldots = b_n = 0$.

Since c_k, $d_k \in R$, it follows that $s(x), p(x) \in R[x]$; thus closure properties hold in $R[x]$, with respect to both addition and multiplication. The associative properties of addition and multiplication, the commutative property of addition, and the two distributive properties can readily be verified in $R[x]$. The null polynomial is the null element of $R[x]$; and the negative of a polynomial $a_0 + a_1 x + \ldots + a_n x^n$ is given by the polynomial $(-a_0) + (-a_1)x + \ldots + (-a_n)x^n \in R[x]$.

If R be in particular, a commutative ring, the commutative property of multiplication holds also in $R[x]$.

Let R be an integral domain. Let $f(x)$ and $g(x)$ be non-null polynomials; and let r and s be the smallest suffices such that $a_r \neq 0$ and $b_s \neq 0$. Then, for the product polynomial $p(x)$,

$$\begin{aligned} d_{r+s} &= a_0 b_{r+s} + \ldots + a_{r-1} b_{s+1} + a_r b_s + a_{r+1} b_{s-1} + \ldots + a_{r+s} b_0 \\ &= a_r b_s \text{ (since } a_0 = \ldots = a_{r-1} = 0,\ b_{s-1} = \ldots = b_0 = 0) \\ &\neq 0, \text{ (since } a_r \neq 0,\ b_s \neq 0, \text{ and } R \text{ has no divisors of zero).} \end{aligned}$$

Consequently, $R[x]$ has no divisors of zero; $R[x]$ is therefore an integral domain. Thus one obtains the following:

I.8.1 (i) *The set of all polynomials, over a ring R, forms a ring $R[x]$, called the poynomial ring over R.*

(ii) *If R be a commutative ring, then so also is $R[x]$.*

(iii) *If R be an integral domain, then $R[x]$ is also an integral domain.*

If R be a field, then $R[x]$ is *not* a field, since the polynomial $x+a$ (where $a \varepsilon R$) has no inverse in $R[x]$. But

I.8.2. *If K be a field, then $K[x]$ is a Euclidean domain.*

Proof: Since a field K is always an integral domaim containing identity, so also is $K[x]$, by I.8.1. Let a norm function be defined for the non-null elements of $K[x]$ by

$$[f(x)]N = \text{degree of } f(x)+1.$$

The norm is then a positive integer. Also,

$$\begin{aligned}[f(x)g(x)]N &= \text{degree of } f(x).g(x)+1\\ &= \text{degree of } f(x)+\text{degree of } g(x)+1\\ &= [f(x)]N+[g(x)]N-1 \geqslant f(x)N, \quad \because \; [g(x)]N \geqslant 1.\end{aligned}$$

Thus the norm function satisfies the requisite condition E(1).

Next, given two polynomials,

$$a(x)=a_nx^n+a_{n-1}x^{n-1}+\ldots+a_1x+a_0$$

and
$$b(x)=b_mx^m+b_{m-1}x^{m-1}+\ldots+b_1x+b_0,$$

where $a_n \neq 0$, $b_m \neq 0$, and $n \geqslant m$, two polynomials

$$c(x)=c_{n-m}x^{n-m}+\ldots+c_1x+c_0$$

and
$$d(x)=d_{m-1}x^{m-1}+\ldots+d_1x+d_0$$

are uniquely determined, such that

$$a(x)=b(x)\,c(x)+d(x). \qquad \ldots (1)$$

In fact, the coefficients of $c(x)$ and $d(x)$ are determined successively by equating the coefficients of like powers of x from the two sides of (1). Also, $[d(x)]N < m \leqslant [b(x)]N$. Hence division algorthm, satisfying E(2), is defined in $R[x]$. $R[x]$ is therefore a Euclidean domain.

It can be proved that *if D be a Gaussian domain, then $D[x]$ is also a Gaussian domain.* But if D be a Euclidean domain, then $D[x]$ may *not* be a Euclidean domain, nor even a principal ideal domain.

Roots of a polynomial. Let $f(x)=a_nx^n+\ldots+a_1x+a_0$ be a polynomial of degree n, over a field K. An element u, belonging to K or to an extension M of K, is said to be a *root of* $f(x)$, if $f(u)=a_nu^n+\ldots+a_1u+a_0=0$. In $M[x]$, which is a Euclidean domain, $f(x)$ can be expressed as $f(x)=(x-u).q(x)+r$. Then $0=f(u)=r$, so that $x-u$ is a factor of $f(x)$ in $M[x]$. And conversely, if $x-u$ is a factor of $f(x)$, then u is a root of $f(x)$. Thus *u is a root of $f(x)$ in M if, and only if, $x-u$ is a factor of $f(x)$ in $M[x]$.*

Let u be a root of $f(x)$ in M. Then there exists a positive integer r, such that $(x-u)^r$ is a factor of $f(x)$, but $(x-u)^{r+1}$ is not a factor of $f(x)$, in $M[x]$; r is called the *order* or *the multiplicity of the root u.* A root of multiplicity 1 is called a simple root, while a root whose multiplicity is greater than 1 is called a *multiple root.* An irreducible polynomial $f(x)$, every root of which is simple, is called a *separable polynomial* ; otherwise, the polynomial is said to be *non-separable.* It can be shown that *if $f(x)$ be an irreducible polynomial over a field K then $f(x)$ is separable, unless K has characteristic p and $f(x)$ is expressible as a polynomial in x^p, i.e., $f(x)=g(x)^p$, over K.* In particular, *an irreducible polynomial over a field of characteristic 0, or over a Galois field, is always separable.*

Existence of a root of any given polynomial, over a field, has been proved by Kronecker in the following theorem :

I.8.3. Fundamental theorem of general algebra : *If $f(x)$ be a polynomial of degree $\geqslant 1$, over a field K, then there exists an extension M of K, such that M contains a root of $f(x)$.*

It can then be shown that there exists an an extension W of K, such than $f(x)$ decomposes into linear factors,

$$f(x)=a_n(x-u_1)\ldots(x-u_n) \text{ in } W[x].$$

The smallest extension W, with the above property, is uniquely determined ; it is called the *splitting field* or *root field of $f(x)$ over K.*

9. Algebraic extensions of a field.

Let M be an extension of a field K. An element u of M is said to be *algebraic over K*, if u is a root of a (non-null) polynomial $f(x) \varepsilon K[x]$; otherwise, u is said to be *transcendental over K*.

Let u be algebraic over a field K, being a root of a polynomial $f(x)$ of degree n, which belongs to $K[x]$, and is irreducible in $K[x]$. Then the set of all elements.

$$a_0+a_1u+\ldots\ldots+a_{n-1}u^{n-1}, \text{ where } a_0, a_1, \ldots, a_{n-1} \varepsilon K,$$

forms an extension of K, called the *simple extension* $K(u)$; u is called a *primitive element* of the extension. It can be shown that every element of $K(u)$ is algebraic over K.

If, in particular, $n=2$, the simple extension $K(u)$ is called a *quadratic extension*. If u' be the other root of the quadratic $f(x)$, then $K(u')=K(u)$; and the two elements $a+bu$, $a+bu'$, where $a, b \varepsilon K$, are said to be conjugate of one another, over K. E.g. the field of complex numbers C is a quadratic extension of the real field E, $C=E(i)$; $a+bi$ and $a-bi$ are conjugates over E.

M is said to be an *algebraic extension of K*, if every element of M is algebraic over K. Any simple extension $K(u)$, where u is algebraic over K, is an algebraic extension, but not conversely.

A field F is said to be *algebraically closed* (or, simply, *closed*), if there does not exist any algebraic extension of F. A necessary and sufficient condition for a field F to be algebraically closed is that every polynomial, over F, decomposes into linear factors in $F[x.]$ Existence of an algebraically closed field has been proved by Gauss in the following theorem :

I.9.1. Fundamental theorm of classical algebra : *The field of complex numbers is algebraically closed.*

There exist other fields which are also algebraically closed. E.g. the algebraic numbers form an algebraically closed field. In fact, Steinitz has proved that *for any given field K, there exists an extension of K which is algebraically closed*; it is called the *algebraic closure of K*.

A field K is said to be *formally real*, if -1 cannot be expressed as a square or as a sum of squares of elements of K. The characteristic of a formally real field is necessarily 0. A subfield of a formally real field is evidently formally real.

A formally real field K is said to be *R-closed*, if there does not exist any formally real proper algebraic extension of K. A necessary and sufficient condition that a field K be R-closed is that the set of all elements $a+bi$, where $a, b \varepsilon K$ and i is a root of x^2+1 (over K), forms an algebraically closed field. It then follows immediately from the fundamental theorem of classical algebra that *the field of real numbers is R closed*.

10. Remarks.

Some definitions and theorems of abstract algebra have been stated above, with outlines of proofs in some cases. For detailed proofs and a connected account of the above concepts, one should read a book on abstract algebra. These definitions and theorems will be referred to in the following chapters.

In developing the theory of linear spaces and matrices, one could as well restrict the ground system (i.e., the system of *scalars*) to the field of complex numbers. In fact, for the purpose of application of the results of linear algebra to geometries, physics and other sciences, the field of complex numbers is found to be sufficient, in most of these cases ; but more abstract scalar systems are sometimes found to be neccssary.

The theory of linear spaces is being developed, in the following chapters, over an abstract field (instead of the field of complex numbers, in particular). This does not make the subject complicated ; rather it simplifies matters in most of the cases.

Almost all the results proved in the following chapters remain valid, if one considers (*i*) the field of complex numbers in place of any abstract skew field or an algebraically closed field, (*ii*) the ring of integers in place of an abstract integral domain or a Gaussian domain or a Euclidean domain, and (*iii*) the field of real numbers in place of an abstract formally real field.

CHAPTER II

LINEAR SPACES

1. Linear space and sub-spaces.

Let V be an additive commutative group and S a skew-field; their elements will ordinarily be represented respectively by Greek and Roman letters, with or without any suffix. A combination of elements of S and V is defined so that any $c \varepsilon S$ and $\alpha \varepsilon V$ determine a unique element in V, which is denoted by $c.\alpha$ or $c\alpha$. This combination is subject to the following conditions:

$L(1)$: $1.\alpha = \alpha$; $L(2)$: $c.(d\alpha) = (cd).\alpha$;

$L(2)$: $(c+d).\alpha = (c.\alpha) + (d.\alpha)$; $L(4)$: $c.(\alpha+\beta) = (c.\alpha) + (c.\beta)$.

V is then said to be a *linear space* or *vector space* over the skew-field S. Elements of V and S are respectively called *vectors* and *scalars*, and the combination of vector and a scalar is called *multiplication of a vector by a scalar.* The null element of V is called the *null vector*; it will be denoted by θ. The null and the identity elements of S, i.e., 0 and 1, are called the *null scalar* and the *identity scalar.*

II.1.1. (i) $0\alpha = \theta$, (ii) $(-1).\alpha = -\alpha$.

Proof: Let c be any scalar, then from $c.\alpha = (c+0)\alpha = (c.\alpha) + (0.\alpha)$, it follows that $0\alpha = \theta$. Next

$$\theta = 0.\alpha = [1 + (-1)].\alpha = (1.\alpha) + [(-1).\alpha] = \alpha + (-1).\alpha;$$

hence $(-1).\alpha = -\alpha$.

A subset T of a linear space V is called a *linear sub-space* or a *sub vector space*, or simply a *sub-space* of V, if T is by itself a linear space, over S, with respect to the same compositions of addition of vectors and multiplication of vectors by scalars, as defined in V.

In what follows, V shall denote a linear space, over a given skew field S.

II.1.2. *A subset T of V is a sub-space of V if, and only if, the sum of two vectors of T lies in T and the product of any vector of T by any scalar lies in T.*

Proof: Let $\alpha, \beta \varepsilon T$, then $-\beta = (-1).\beta \varepsilon T$; and so the difference $\alpha - \beta = \alpha + (-\beta) \varepsilon T$. Hence by I.4 3, T is a subgroup of V, so that T is by itself an additive commutative group. $L(1)-L(4)$ are also satisfied in T, since these properties are hereditary. Thus T is by itself a linear space, over S; i.e., T is a sub-space of V. The converse is an immediate consequence of the definition of a sub-space.

II.1.3. *The intersection of any aggregate of sub-spaces of V is a sub-space of V.*

Proof: Let $M = \cap [T$ be the intersection of a given aggregate of sub-spaces $[T]$ of V. Let $\alpha, \beta \varepsilon M$ and $a \varepsilon S$. Then α, β belong to *every* T; and since each T is a sub-space, $\alpha - \beta$ and $a.\alpha$ belong to each T, by II.1 2. Consequently $\alpha - \beta$, $a.\alpha \varepsilon M$; hence M is a sub-space of V, by II.1.2.

$\cap [T]$ is evidently the largest sub-space of V contained in all the sub-spaces of the given aggregate $[T]$.

The join $P \cup Q$ of two sub-spaces is not necessarily a sub-space. For instance, if P consists of all multiples of a vector α and Q consists of all multiples of β, where β does not belong to P, then $P \cup Q$ does not contain the sum $\alpha + \beta$; hence $P \cup Q$ cannot be a sub-space of V.

The set of all vectors $\sigma + \tau$, where σ and τ run over two given sub-spaces P and Q of a linear space V is linear sub-space $P+Q$ (by II.1.4), called the *linear sum* of P and Q. Any sub-space containing the two given sub-spaces P and Q must contain every sum $\sigma + \tau$. Hence the linear sum $P+Q$ is the smallest sub-space containing both the sub-spaces P and Q.

In general, the *linear sum* of a given aggregate of sub-spaces $[T]$ is defined, to be the subset W, consisting of all those elements of V, which are fiinite sums of elements taken from the sub-spaces belonging to the aggregate $[T]$.

II.1.4. *The linear sum of any aggregate of sub-spaces of V is a sub-space of V.*

Proof: Let W be the linear sum of a given aggregate of sub-spaces $[T]$ of V. Then, for any two elements λ, μ of W, there exists a finite number of T's, denoted by $T_1, \ldots T_n$, say, such that

$$\lambda = a_1\alpha_1 + \ldots + a_n\alpha_n \text{ and } \mu = b_1\alpha_1 + \ldots + b_n\alpha_n,$$

where $\alpha_i \varepsilon T_i$ and $a_i, b_i \varepsilon S$, for $i = 1, 2, \ldots, n$ (some of the a's or b's may also be null). Then,

$$\lambda + \mu = (a_1 + b_1)\alpha_1 + \ldots + (a_n + b_n)\alpha_n \varepsilon W.$$

and $$c\lambda = (ca_1)\alpha_1 + \ldots + (ca_n)\alpha_n \varepsilon W, \text{ for any } c \varepsilon S.$$

Hence W is, by II.1.2, a sub-space of V.

Generators. Let B be a subset of V. Then the intersection of all sub-spaces containing B is a sub-space T of V. It is called the sub-space *generated* or *spanned* by the subset B; B is called *a set of generators* of T.

II.1.5. *The sub-space T, generated by a subset B of a linear space V, consists of all linear combinations,*

$$c_1\beta_1 + \ldots + c_n\beta_n, \qquad \ldots(1)$$

for all finite sub-sets of vectors $\beta_1, \ldots, \beta_n$ belonging to B, where $c_1, \ldots, c_n$ run over S.

Proof: Let P consist of all linear combinations (1). Then the sum of two linear combinations,

$$(b_1\beta_1 + \ldots + b_n\beta_n) + (c_1\beta_1 + \ldots + c_n\beta_n)$$
$$= (b_1 + c_1)\beta_1 + \ldots + (b_n + c_n)\beta_n \varepsilon P;$$

and the product of a linear combination by a scalar,

$$b.(c_1\beta_1 + \ldots + c_n\beta_n) = (bc_1)\beta_1 + \ldots + (bc_n)\beta_n \varepsilon P.$$

Hence P is a sub-space of V, by II.1.2. Also, since P contains B, $T \subset P$ (as T is the intersection of all sub-spaces like P).

Again, since T is a linear sub-space, containing B, it contains all linear combinations (1); consequently, $P \subset T$. Combining the results, $T \subset P$ and $P \subset T$, one gets $T = P$.

In particular, the sub-space generated by a single vector β consists of all multiples $c.\beta$; the sub-space generated by two vectors α and β consists of all vectors $c\alpha + d\beta$, where c, d run over K.

It now follows from II.1.5 that

II.1.6. *A subset B of a linear space V forms a set of generators of V, if every vector belonging to V can be expressed as a linear combination of a finite number of vectors from B.*

Also it follows readily from II.1.5, that ;

II.1.7. *If B be a subset of V and if $C \subset B$, then the sub-space C^* generated by C is a sub-space of the sub-space B^* generated by B.*

It should however be noted that when C is a proper subset of B, C^* may even be identical with B^* (see II.1.10).

II.1.8. Replacement theorem. *Let T be a sub-space generated by a subset B. Let B' be obtained from B by replacing a vector α_i by a linear combination $c_1\alpha_1+\dots+c_m\alpha_m$, where $\alpha_1,\dots,\alpha_m \,\varepsilon\, B$, $c_i \neq 0$ and $i \leqslant m$, then the set B' also generates T.*

Proof: Let B' generate the linear space T', then it follows readily from II.1.5 that $T' \subset T$. Again, any vector of T', which is a linear combination of vectors belonging to B, but not involving α_i, certainly belongs to T'. Let $c_1\alpha_1+\dots+c_m\alpha_m=\lambda$, then any linear combination of vectors of B, involving α_i,

$$
\begin{aligned}
& a_1\alpha_1+\dots+a_m\alpha_m \\
&= a_1\alpha_1+\dots+a_{i-1}\alpha_{i-1}+a_{i+1}\alpha_{i+1}+\dots+a_m\alpha_m \\
&\qquad +(c_i^{-1}a_i)[\lambda-c_1\alpha_1-\dots-c_{i-1}\alpha_{i-1}-c_{i+1}\alpha_{i+1}-\dots-c_m\alpha^{m}] \\
& \varepsilon\, T',
\end{aligned}
$$

so that $T \subset T'$. Hence $T'=T$.

Corollary. *A linear space T possesses different sets of generators.*

1 ✓ **Linear dependence.** A vector α is said to be *linearly dependent on* a given set of vectors $B \subset V$, if α belongs to the sub-space generated by B, i.e., if α can be expressed as a linear combination of a finite number of vectors belonging to B.

A given set of vectors B is said to be *linearly independent* if, for any set of vectors $\alpha_1, \alpha_2, \dots, \alpha_r \,\varepsilon\, B$,

$$c_1\alpha_1+\dots+c_r\alpha_r=\theta,$$

only when $c_1=\dots=c_r=0$; otherwise B is said to be a *linearly dependent set.*

II.1.9. *A set of vectors $B \subset V$ is linearly dependent if, and only if, a suitable vector $\alpha \varepsilon B$ is linearly dependent on the set of the remaining vectors in B.*

Proof: Let B be a linearly dependent set, then there exists a finite number of vectors $\alpha_1, \ldots, \alpha_r \varepsilon B$, such that $a_1\alpha_1 + \ldots + a_r\alpha_r = \theta$, where the scalars $a_1, \ldots, a_r$ are not all null. If $a_k \neq 0$ $(1 \leq k \leq r)$, then

$$\alpha_k = b_1\alpha_1 + \ldots + b_{k-1}\alpha_{k-1} + b_{k+1}\alpha_{k+1} + \ldots + b_r\alpha_r,$$

where $b_i = -a_k^{-1}.a_i$, for $i = 1, \ldots, k-1, k+1, \ldots, r$. Thus α_k is linearly dependent on the set of vectors $\alpha_1, \ldots, \alpha_{k-1}, \alpha_{k+1}, \ldots, \alpha_r$, and hence also on the set of all vectors in B excluding α_k.

The converse follows by reversing the above arguments.

II.1.10. *A set of vectors B is linearly dependent if, and only if, a proper subset of B generates the same sub-space as B.*

Proof: Let B be a linearly dependent set, then there exists a vector $\alpha \varepsilon B$ such that $\alpha = c_1\alpha_1 + \ldots + c_m\alpha_m$, where $\alpha_1, \ldots, \alpha_m \varepsilon B$ and the scalars $c_1, \ldots, c_m$ are not all null. If, now, B' be the set of vectors obtained from B by replacing the vector α by $\alpha - c_1\alpha_1 - \ldots - c_m\alpha_m = \theta$, then B', i.e., the non-null vectors in B', generates the same sub-space as B (by II.1.8); also, the non-null vectors in B' form a proper subset of B.

Conversely, if C be a proper subset of B, such that B and C generate the same sub-space and if $\beta \varepsilon B - C$, then β is linearly dependent on C, and hence also on the set of all vectors of B excluding β. Consequently, B is a linearly dependent set (by II.1.9.).

For a finite set B, proposition II.1.9. may be proved in the following particular form:

II.1.11. *An ordered set of non-null vectors in V, finite in number, is linearly dependent if, and only if, some one of the vectors be linearly dependent on the preceding vectors.*

Proof: Let $\alpha_1, \ldots, \alpha_n$ be a given set of vectors in V. If the vector α_k be linearly dependent on the preceding vectors, say $\alpha_k = c_1\alpha_1 + \ldots + c_{k-1}\alpha_{k-1}$, then the linear combination,

$$c_1\alpha_1 + \ldots + c_{k-1}\alpha_{k-1} + (-1)\alpha_k + 0.\alpha_{k+1} + \ldots + 0.\alpha_n = \theta\ ;$$

hence the given set is linearly dependent.

Conversely, if the given set of vectors be linearly dependent, there exist suitable scalars, $b_1,\ldots,b_n$, not all all of which are null, such that the linear combination $b_1\alpha_1+\ldots+b_n\alpha_n=0$. Let k be the suffix for which $b_k\neq 0$, but $b_{k+i}=0$ for $i=1,\ldots,n-k$, then one can express α_k as

$$\alpha_k=(-b_k^{-1}.b_1).\alpha_1+\ldots+(-b_k^{-1}.b_{k-1}).\alpha_{k-1},$$

which is a linear combination of vectors preceding α_k.

Dimensionality : A linear space V is said to be *finite dimensional*, if there exists a finite set of vectors generating V; otherwise V is said to be *infinite dimensional.*

The concept of dimension of a linear space in a more concrete form will be introduced in the next article.

EXAMPLE (1) Let the aggregate of all n-tuples (ordered set of n elements) with elements from a skew-field S be denoted by $V_n(S)$. Let addition of two n-tuples and multiplication of an n-tuple by an element of S be defined by :

$$(a_1,\ldots,a_n)+(b_1,\ldots,b_n)=(a_1+b_1,\ldots,a_n+b_n)$$

and
$$c.(a_1,\ldots,a_n)=(ca_1,\ldots,ca_n),$$

then $V_n(S)$ becomes a linear space over S. The properties of linear space may easily be verified for $V_n(S)$, $(0,0,\ldots 0)$ being the null vector and the negative of an n-tuple $(a_1,\ldots.a_n)$ being given by $(-a_1,\ldots,-a_n)$.

This linear space is finite dimensional, since the n-tuples $(1,0,\ldots,0)$, $(0,1,\ldots,0),\ldots,(0,0,\ldots.1)$ generate this space. In fact, any arbitary n-tuple can always be expressed as a linear combination of these n-tuples :

$$(a_1,a_2\ldots,a_n)=a_1(1,\ 0,\ldots,0)+a_2(0,1,\ldots.0)+\ldots+a_n(0,0.\ldots,1).$$

(2) The set of all polynomials, $a_0+a_1x+a_2x^2+\ldots+a_mx^m$, for all finite values of m, where the coefficients a_i belong to a skew field K, forms a linear space $K[x]$ over K, the addition of polynomials and the multiplication of a polynomial by a scalar being defined in the usual way.

This linear space is infinite dimensional, since any linear combination of a finite number of polynomials will be a polynomial whose degree does not exceed the maximum degree of the polynomials in the given set ; but the linear space contains a polynomial of degree as large as one likes.

Exercises

1. All coplaner forces acting on a particle form a linear space F_2 over the field of real numbers. F_2 is finite dimensional.

2. (i) All twice-differentiable functions $f(x)$ of a real variable x form a linear space D_2, over the field of real numbers.

(ii) All solutions of the linear homogeneous differential equation

$$\frac{d^2f(x)}{dx^2}-5\frac{df(x)}{dx}+6f(x)=0$$

form a sub-space H of D_2.

(iii) The vector space D_2 is infinite dimensional; but the sub-space H is finite dimensional.

3. (i) All homogeneous polynomials of degree m in two variables x, y, with coefficients from a skew field K, form a linear space $K_m[x, y]$ over K.

(ii) The symmetric polynomials in $K_m[x, y]$ (viz. those which remain unaltered by interchanging x and y) form a sub-space of $K_m[x, y]$.

4. The set of all solutions $(x_1, \ldots, x_n)$ of a system of linear homogeneous equations

$$\left.\begin{array}{l} a_{11}x_1+a_{12}x_2+\ldots+a_{1n}x_n=0 \\ a_{21}x_1+a_{22}x_2+\ldots+a_{2n}x_n=0 \\ \ldots \quad \ldots \quad \ldots \quad \ldots \\ a_{m1}x_1+a_{m2}x_2+\ldots+a_{mn}x_n=0 \end{array}\right\},$$

with coefficients from a skew field S, is a subspace of $V_n(S)$.

5. Determine which of the following are sub-spaces of V_n(E), where E is the field of real numbers:

The set of all n-tuples $(a_1, a_2, \ldots, a_n)$ for which

(i) $a_1=0$,
(ii) $a_1=$an integer,
(iii) $a_1a_2=0$,
(iv) $2a_1+5a_2=0$,
(v) $2a_1-3a_2=1$,
(vi) $a_1=a_2=2a_3$,
(vii) $a_1^2=a_2a_3$,
(viii) $a_1=a_2=a_3+a_4$,

[where $n\geqslant 2$ for (iii)-(v), $n\geqslant 3$ for (vi), (vii), and $n\geqslant 4$ for (viii)].

Taking $n=5$, find the intersection and linear sum of every pair of *sub-spaces* (i)-(viii).

6. Let T be the sub-space of a linear space V, generated by a linearly independent set of vectors $\alpha, \beta, \gamma, \delta$, then

(i) $\alpha+\beta+\gamma+\delta$, $\alpha+\beta+\gamma$, $\alpha+\beta$, α generate T,

(ii) $\alpha-\beta$, $\alpha-\gamma$, $\alpha-\delta$, $\beta-\gamma$, $\beta-\delta$, $\gamma+\delta$ generate T,

(iii) $\alpha+\beta$, $\alpha+\gamma$, $\alpha+\delta$ do not generate T.

2. Finite dimensional linear spaces.

Let V be a linear space over a skew field S.

R **Basis.** A linearly independent set of vectors generating the linear space V is said to form a *basis* of V.

R II.2.1. *There exists a basis for every finite dimensional linear space V.*

Proof: Since V is finite dimensional, there exists a finite set of generators $\gamma_1, \ldots, \gamma_m$ of V. If this set be not linearly independent, then by II.1.9, one can omit some suitable vector from this set such that the remaining vectors also generate V. If the remaining vectors be linearly independent, then they form a basis of V, and the theorem is proved; otherwise a suitable vector is again omitted from this reduced set of generators, and this process is repeated till a linearly independent set of generators of V, i.e., a basis of V is obtained. The set of generators at the initial stage being finite, the basis is obtained in a finite number of steps.

Let V be finite dimensional. Since the vectors, constituting a basis, form a linearly independent set, it follows, from II.1.10, that if a vector be omitted from a basis the remaining vectors generate only a proper sub-space of V. Thus a basis gives an irreducible set of vectors generating the space V.

76/1 II.2.2. *Any linearly independent set of vectors belonging to V can always be extended to a basis of V (in case it is not already a basis).*

Proof: Let $\gamma_1, \ldots, \gamma_m$ be a set of generators of V, and let $\alpha_1, \ldots, \alpha_r$ be a linearly independent set of vectors in V. The vectors $\alpha_1,\ldots,\alpha_r$, $\gamma_1,\ldots,\gamma_m$ certainly form a set of generators of V. One can extract, by the method outlined in II.2.1, a basis of V from this set of generators, by deleting, one by one, vectors which are linear combinations of the *preceding* vectors. Since the vectors α_i are linearly independent and occur at the beginning, none of these will be deleted, and so the resulting basis will contain all these given vectors.

By the method, used in II.2.2, a basis is obtained by adding to a given linearly independent set of vectors some suitable vectors from *any* set of generators. In case, the set of generators $\gamma_1,\ldots,\gamma_m$ be itself a basis, and the given vectors α_r, $\alpha_{r-1},\ldots,\alpha_1$ be absorbed one by one, and the extended set of generators be reduced to a basis at every step (that is, at the first step the set α_r, $\gamma_1,\ldots,\gamma_m$ is reduced to a basis, then at the second step α_{r-1} is added to this basis and then reduced to a new basis, and so on) the number of vectors in the resulting basis is greater than or equal to r, and at the same time it is also less than or equal to m. Thus one obtains :

II.2.3. *The number of vectors in any linearly independent set of vectors belonging to V is less than or equal to the number of vectors in any basis of V.*

R**Dimension or rank.** For a given linear space V, one may find by the replacement theorem, many different sets of generators, starting from a particular set of generators. Again by the method outlined in II.2.1, one can extract a basis from every set of generators of V. Hence *there exist different bases for a given linear space V ; but*

P— II.2.4. *Any two bases of a finite dimensional linear space contain the same number of vectors.*

Proof : If V has two bases containing m and n vectors, then each of the bases being a linearly independent set, it follows from II.2.3 that $m \leqslant n$, and $n \leqslant m$. Hence $m = n$.

Thus the number of vectors in a basis of a linear space V appertains to V ; it is called the *rank* or *dimension* of V and will be denoted by $d(V)$.

EXAMPLE 1: The rank of $V_n(S)$ (example 1, page 31) is n, since the vectors $(1, 0,\ldots,0), (0, 1,\ldots,0),\ldots, (0, 0,\ldots, 1)$, which generate $V_n(S)$, form a basis. In fact

$$a_1(1, 0,\ldots, 0)+a_2(0, 1,\ldots, 0)+\ldots+a_n(0, 0,\ldots,1)=(0, 0,\ldots,0),$$

implies $a_1=0,\ a_2=0,\ldots,\ a_n=0$.

II.2.5. *If P and Q are any two sub-spaces of V, then*

$$d(P)+d(Q)=d(P\cap Q)+d(P+Q).$$

Proof: Let $\gamma_1,\ldots,\gamma_n$ form a basis of $P\cap Q$. By II.2.2, one can obtain for P and Q bases $\gamma_1,\ldots,\gamma_n,\ \delta_1,\ldots,\delta_r$ and $\gamma_1,\ldots,\gamma_n,\ \beta_1,\ldots,\beta_k$ respectively. Then

$$d(P)=n+r,\quad d(Q)=n+k,\ \text{and}\ \ d(P\cap Q)=n.$$

Evidently $\gamma_1,\ldots,\gamma_n,\ \delta_1,\ldots,\delta_r,\ \beta_1,\ldots,\beta_k$ generate $P+Q$. Next, let

$$a_1\gamma_1+\ldots+a_n\gamma_n+b_1\delta_1+\ldots+b_r\delta_r+c_1\beta_1+\ldots+c_k\beta_k=\theta.$$

Then $\lambda=\Sigma b_i\delta_i=-\Sigma a_i\gamma_i-\Sigma c_i\beta_i$ lies in both P and Q; hence $\lambda\in P\cap Q$. Consequently, $\Sigma b_i\delta_i=\lambda=\Sigma d_i\gamma_i$, for some scalars d_i. But the vectors $\gamma_1,\ldots,\gamma_n,\ \delta_1,\ldots,\delta_r$ are linearly independent (since these form a basis of P); hence every $b_i=0$ and every $d_i=0$, so that $\lambda=\theta$. Then, every $a_i=0$ and every $c_i=0$ (since $\gamma_1,\ldots,\gamma_n,\ \beta_1,\ldots,\beta_k$ are linearly independent). This shows that the vectors $\gamma_1,\ldots,\gamma_n,\ \delta_1,\ldots,\delta_r,\ \beta_1,\ldots,\beta_k$ are also linearly independent. So they form a basis of $P+Q$, and therefore $d(P+Q)=n+r+k$. Hence the proposition holds.

Complements. A sub-space Q is said to be a *complement* of a given sub-space P in V if $P+Q=V$ and $P\cap Q=\theta$ (i.e., consists of the null vector θ only).

II.2.6. *Every sub-space P has a complement in V.*

Proof: Let the vectors $\sigma_1,\ldots,\sigma_r$ constitute a basis of P; then they form a linearly independent set in V, which can (by II.2.2) be extended to a basis, $\sigma_1,\ldots,\sigma_r,\ \sigma_{r+1},\ldots,\sigma_n$ of V. Then $\sigma_{r+1},\ldots,\sigma_n$ generate a sub-space Q, which is evidently a complement of P in V.

Complement of P in V is *not* uniquely determined (see example 2 (iii), page 39).

Direct sum. If two sub-spaces P and Q are complements of each other in V, then V is said to be the *direct sum* of P and Q, and the relation is denoted by $V=P\oplus Q$. [In fact, V is additively a commutative group having P and Q as two normal subgroups; then by the given conditions, it follows that the additive group V is the direct sum of the two subgroups P and Q.]

A finite set of sub-spaces $T_1,\ldots, T_m$ of V is said to be *independent*, if

$$(T_1+\ldots+T_{k-1}+T_{k+1}+\ldots+T_m)\cap T_k=\theta,\ \text{for } k=1,\ldots,m.$$

The linear sum of a set of independent sub-spaces $T_1,\ldots, T_m$ is called their *direct sum*, and is denoted by $T_1\oplus\ldots\oplus T_m$.

II.2.7. *If V be the direct sum of the sub-spaces $T_1,\ldots, T_m$, then*

(i) every vector ξ belonging to V can be expressed uniquely as $\xi=\sigma_1+\ldots+\sigma_m$, where $\sigma_i\ \varepsilon\ T_i$, for $i=1,\ldots, m$;

(ii) a set of vectors, formed by taking together a basis of T_1, a basis of $T_2,\ldots$, and a basis of T_m, constitutes a basis of V;

(iii) $d(V)=d(T_1)+\ldots+d(T_m)$.

Proof: (i) Since V is the linear sum of the sub-spaces $T_1,\ldots, T_m$, every vector $\xi\ \varepsilon\ V$ can be expressed as

$$\xi=\sigma_1+\ldots+\sigma_m, \text{ where } \sigma_i\ \varepsilon\ T_i, \text{ for } i=1,\ldots, m.$$

If possible, let $\xi=\sigma_1'+\ldots+\sigma_m'$ be a different representation, with $\sigma_k\neq\sigma_k'$. Then

$$\sigma_k-\sigma_k'=(\sigma_1'-\sigma_1)+\ldots+(\sigma_{k-1}'-\sigma_{k-1})+(\sigma_{k+1}'-\sigma_{k+1})\ldots +(\sigma_m'-\sigma_m),$$

so that $(T_1+\ldots+T_{k-1}+T_{k+1}+\ldots+T_m)\cap T_k\supset(\sigma_k-\sigma_k')\neq\theta$, which is impossible (as the sub-spaces $T_1,\ldots, T_m$ are independent).

(ii) Let a basis of T_i, for $i=1,\ldots, m$, be given by

$$\alpha_{i1},\ \alpha_{i2},\ldots,\ \alpha_{ir_i}. \qquad \ldots \qquad (1)$$

Then it follows, by (i), that the collection of all sets of vectors (1), for $i=1,\ldots, m$, is linearly independent. Also, as V is the linear sum of $T_1,\ldots, T_m$, these vectors form a set of generators of V. Hence these vectors form a basis of V.

(iii) This result follows immediately from (ii).

Co-ordinatisation. A biuniform mapping m of a linear space V onto a linear space V', over the same skew field S, is said to be an *isomorphism* if, for any vectors $\alpha, \beta \varepsilon V$ and any scalar $c \varepsilon K$,

$$(\alpha+\beta)\mathrm{m}=\alpha\mathrm{m}+\beta\mathrm{m} \text{ and } (c\alpha)\mathrm{m}=c(\alpha\mathrm{m}).$$

Two isomorphic linear spaces are abstractly equal.

All n-tuples $(x_1,\ldots,x_n)$, where the co-ordinates x run over S, form a linear space $V_n(S)$ of rank n (example 1, page 31). One can now show that an arbitrary linear space V of rank n, over the skew field S, has a representation by the space $V_n(S)$.

II.2.8. *Every vector in V can be expressed as a linear combination of the vectors belonging to a given basis in one and only one way.*

Proof: Let $\mu_1,\ldots,\mu_n$ form a basis of V; then any vector ξ belonging to V can always be expressed as a linear combination $\xi=x_1\mu_1+\ldots+x_n\mu_n$. If ξ be equal to a second such linear combination say $\xi=y_1\mu_1+\ldots+y_n\mu_n$, then by taking the difference of these two expressions.

$$(x_1-y_1)\mu_1+\ldots+(x_n-y_n)\mu_n=\xi-\xi=\theta.$$

Since the basis vectors are linearly independent, each coefficient of this linear combination must be null, so the $x_i=y_i$, for $i=1, \ldots, n$.

II.2.9. *Any linear space V of rank n, over a skew field S, is isomorphic with the (concrete) linear space $V_n(S)$.*

Proof: Let $\mu_1,\ldots,\mu_n$ be a basis of V, then by II.2.8, every vector of V can be expressed uniquely as a linear combination $\xi=x_1\mu_1+\ldots+x_n\mu_n$. Then ξ determines a unique n-tuple $(x_1,\ldots,x_n) \varepsilon V_n(S)$. Thus $\xi\mathrm{m}=(x_1,\ldots,x_n)$ is a biuniform mapping of V onto $V_n(S)$. Also, if $\eta\mathrm{m}=(y_1,\ldots,y_n)$, then,

$$\begin{aligned}(\xi+\eta)\mathrm{m}&=[(x_1\mu_1+\ldots+x_n\mu_n)+(y_1\mu_1+\ldots+y_n\mu_n)]\mathrm{m}\\&=[(x_1+y_1)\mu_1+\ldots+(x_n+y_n)\mu_n]\mathrm{m}\\&=(x_1+y_1,\ldots,x_n+y_n)=(x_1,\ldots x_n)+(y_1,\ldots,y_n)\\&=\xi\mathrm{m}+\eta\mathrm{m},\end{aligned}$$

and $$\begin{aligned}(c.\xi)\mathrm{m}&=[c(x_1\mu_1+\ldots+x_n\mu_n)]\mathrm{m}=[(cx_1)\mu_1+\ldots+(cx_n)\mu_n]\mathrm{m}\\&=(cx_1,\ldots,cx_n)=c(x_1,\ldots x_n)=c.(\xi\mathrm{m}).\end{aligned}$$

The n-tuple $(x_1,\ldots,x_n)$ corresponding to a given vector ξ is said to be the *co-ordinate vector* of ξ with respect to the basis $\mu_1,\ldots,\mu_n$. In particular, to the basis vectors $\mu_1, \ldots, \mu_n$ there correspond the co-ordinate vectors $(1,0,\ldots,0)$, $(0,1,\ldots,0),\ldots,(0,0,\ldots,1)$ respectively. The representation is unique so long as one considers a fixed basis of V. Such a representation is called a *co-ordinatisation* of V.

Co-ordinate transformations. The relation between different co-ordinatisations of V, corresponding to different bases, may now be considered.

Let $\alpha_1,\ldots,\alpha_n$ and $\beta_1,\ldots,\beta_n$ be two different bases of a linear space V, of rank n, over a skew field S. Let $\xi \varepsilon V$, then ξ will be represented by two co-ordinate vectors, say $(x_1,\ldots,x_n)$ and $(x_1',\ldots,x_n')$ respectively with respect to the two bases. Since $\beta_1,\ldots,\beta_n$ form a basis of V, the vectors $\alpha_1,\ldots,\alpha_n$ can be expressed in the form

$$\alpha_i = a_{i1}\beta_1 + \ldots + a_{in}\beta_n = \sum_{k=1}^{n} a_{ik}\beta_k \quad \text{for } i=1,\ldots,n, \qquad \ldots (2)$$

where the coefficients a_{ik}, for $i=1,\ldots,n$ and $k=1,\ldots,n$, belong to S, and are uniquely determined by (2) in terms of the two given bases. Now

$$\sum_{k=1}^{n} x_k'\beta_k = \xi = \sum_{i=1}^{n} x_i\alpha_i = \sum_{i=1}^{n} x_i\Big[\sum_{k=1}^{n} a_{ik}\beta_k\Big] = \sum_{k=1}^{n}\Big[\sum_{i=1}^{n} x_i a_{ik}\Big]\beta_k,$$

i.e.,

$$\sum_{k=1}^{n}\Big[x_k' - \sum_{i=1}^{n} x_i a_{ik}\Big]\beta_k = \theta.$$

Since $\beta_1,\ldots,\beta_n$ are linearly independent, it follows that

$$x_k' = \sum_{i=1}^{n} x_i a_{ik}, \quad \text{for } k=1,\ldots,n. \qquad \ldots (3)$$

Thus one obtains:

II.2.10. *A change from a basis $\alpha_1,\ldots,\alpha_n$ to a basis $\beta_1,\ldots,\beta_n$ of a linear space induces a transformation of co-ordinate vectors, represented by the equations* (3), *where the coefficients a_{ik} are uniquely determined by the relation* (2).

EXAMPLE 2 : Consider in V_5 (R), where R is the field of rational numbers, the vectors

$$\kappa=(2,\ -4,\ 5,\ 6,\ -8), \qquad \mu=(-3,\ 6,\ 7,\ -4,\ 12),$$
$$\phi=(-5,\ 10,\ -3,\ 2,\ 4), \qquad \rho=(-6,\ 12,\ 9,\ 4,\ 8),$$

Let T be the sub-space generated by κ, μ, ϕ and ρ. It is required

(i) to find whether κ, μ, ϕ and ρ from a linearly independent set ;

(ii) to find a basis of T, and then the rank of T ;

(iii) to find a complement of T in $V_5(R)$.

To find the linear dependence, or otherwise, let $x\kappa+y\mu+z\phi+t\rho=\theta$, where x, y, z, $t \in R$. Then

$$\begin{aligned}(0,0,0,0,0)&=x(2,\ -4,\ 5,\ 6,\ -8)+y(-3,\ 6,\ 7,\ -4,\ 12)\\ &\quad+z(-5,\ 10,\ -3,\ 2,\ 4)+t(-6,\ 12,\ 9,\ 4,\ 8)\\ &=(2x-3y-5z-6t,\ -4x+6y+10z+12t,\ 5x+7y-3z+9t,\\ &\qquad 6x-4y+2z+4t,\ -8x+12y+4z+8t).\end{aligned}$$

Hence the vectors κ, μ, ϕ and ρ are linearly dependent if, and only if, the following system of linear homogeneous equations

$$\left.\begin{aligned}2x-3y-5z-6t&=0\\ -4x+6y+10z+12t&=0\\ 5x+7y-3z+9t&=0\\ 6x-4y+2z+4t&=0\\ -8x+12y+4z+8t&=0\end{aligned}\right\}$$

has any solution other than $(x,\ y,\ z,\ t)=(0,\ 0,\ 0,\ 0)$. Thus one is led to the problem of solving a system of linear homogenous equations. To avoid this, one can apply the following method, called the *method of sweep out*, by repeated application of the replacement theorem II.1.8 (which, in substance, is equivalent to the method of solving the above system of equations by successive elimination of the unknowns). By this method both the problems (i) and (ii) are solved simultaneously.

(i) By the replacement theorem, it follows that the subspace T is generated by each of the following sets of vectors :

$$\left.\begin{array}{c}\kappa\\ \mu\\ \phi\end{array}\right\}\qquad \left.\begin{array}{rlllll}\kappa_1= & \kappa/2 =(1, & -2, & 5/2, & 3, & 4)\\ \mu_1=\mu+3\kappa/2 & =(0, & 0, & 29/2, & 5, & 0)\\ \phi_1=\phi+5\kappa/2 & =(0, & 0, & 19/2, & 17, & -16)\\ \rho_1=\rho+3\kappa & =(0, & 0. & 24, & 22, & -16)\end{array}\right\},$$

$$\left.\begin{array}{rlllll}\kappa_2=\kappa_1-5\mu_1/29 & =(1, & -2, & 0, & 62/29, & -4)\\ \mu_2= 2\mu_1/29 & =(0, & 0, & 1, & 10/29, & 0)\\ \phi_2=\phi_1-19\mu_1/29 & =(0, & 0, & 0, & 398/29, & -16)\\ \rho_2=\rho_1-48\mu_1/29 & =(0, & 0, & 0, & 398/29, & -16)\end{array}\right\},$$

$$\left.\begin{array}{llll}
\kappa_3=\kappa_2-62\phi_2/393 & =(1, & -2, 0, 0, & -300/199) \\
\mu_3=\mu_2-10\phi_2/898 & =(0, & 0, 1, 0, & 80/199) \\
\phi_3=\quad 29\phi_2/398 & =(0, & 0, 0, 1, & -232/199) \\
\rho_3=\rho_2\quad -\phi_2 & =(0, & 0, 0, 0, & 0)
\end{array}\right\}.$$

Thus $\theta=(0, 0, 0, 0, 0)=\rho_2-\phi_2=(\rho_1-48\mu_1/29)-(\phi_1-19\mu_1/29)$
$=\rho_1-\phi_1-\mu_1=(\rho+8\kappa)-(\phi+5\kappa/2)-(\mu+3\kappa/2)=\rho-\mu-\phi-\kappa.$

Hence κ, μ, ϕ and ρ are linearly dependent.

(ii) The vectors κ_3, μ_3 and ϕ_3 generate T; these are also linearly independent, since $x\kappa_2+y\mu_3+z\phi_3=\theta$ means

$$(x, -2x, y, z, [-800x+80y-232z]/199)=(0, 0, 0, 0, 0),$$

so that $(x, y, z)=(0, 0, 0)$. Thus κ_3, μ_3, ϕ_3 form a basis of T.

Hence the rank of T is 8.

(iii) A basis of $V_5(R)$ is obtained by adding to κ_3, μ_3, ϕ_3 two suitable vectors. The vectors $\varepsilon_1=(1, 0, 0, 0, 0)$, $\varepsilon_2=(0, 1, 0, 0, 0)$, $\varepsilon_3=(0, 0, 1, 0, 0)$, $\varepsilon_4=(0, 0, 0, 1, 0)$ and $\varepsilon_5=(0, 0, 0, 0, 1)$ form a basis of $V_5(K)$, and in this case *any two* of ε_1, ε_2, ε_3, ε_4, ε_5 together with κ_3, μ_3, ϕ_3 will form a basis.

To prove that ε_1, ε_2, κ_3, μ_3, ϕ_3 form a basis, one notes that

$$\varepsilon_5=-199(\kappa_3-\varepsilon_1+2\varepsilon_2)/300.$$
$$\varepsilon_4=\phi_3+282\varepsilon_5/199=\phi_3-58(\kappa_3-\varepsilon_1+2\varepsilon_2)/75,$$
$$\varepsilon_3=\mu_3-80\varepsilon_5/199=\mu_3+4(\kappa_3-\varepsilon_1+2\varepsilon_2)/15,$$

so that an arbitrary vector

$$(a_1, a_2, a_3, a_4, a_5)=a_1\varepsilon_1+a_2\varepsilon_2+a_3\varepsilon_3+a_4\varepsilon_4+a_5\varepsilon_5$$
$$=a_1\varepsilon_1+a_2\varepsilon_2+a_3\mu_3+a_4\phi_3+(80a_3-282a_4-199a_5)(-\varepsilon_1+2\varepsilon_2+\kappa_3)/300.$$

Thus each of the sub-spaces T_{mn}, where T_{mn} is generated by ε_m, ε_n, for $m, n=1, 2, 3, 4, 5$ $(m\neq n)$, is a complement of T in $V_5(R)$. One can obtain other complements by utilising bases other than $(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5)$.

Exercises

[Let R and E denote the field of rational and the field of real numbers respectively].

1. Prove that, if α, β, γ are linearly independent in a linear space V over a skew field S, then so also are the vectors $\alpha+\beta$, $\beta+\gamma$, $\alpha+\gamma$.

2. Show that in a linear space of rank n,

(i) any set of $n+1$ vectors is linearly dependent;

(ii) a set of n vectors forms a basis if, and only if, the set is linearly independent.

3. Show that in an infinite dimensional linear space there exists a set of n linearly independent vectors, for $n=1, 2, \ldots$

4. Test the following sets of vectors for linear dependence :

(i) $(4, 0, 1), (1, 3, 2), (3, 2, 4)$ in $V_3(\mathrm{R})$;
(ii) $((4), (0), (1)), ((1), (3), (2)), ((3), (2), (4))$ in $V_3(GF(5))$;
(iii) $(2, 7, \sqrt{5}), (1, 3, 2), (-\sqrt{2}, 1, 4), (0, 1, 7)$ in $V_3(\mathrm{E})$;
(iv) $(3, 1, 0, 5), (2, 1, 2, 4), (1, 0, 2, 3)$ in $V_4(\mathrm{R})$;
(v) $(1, i, 1+i), (i, -1, 2-i), (0, 0, 3)$ in $V_3(\mathrm{E}(i))$.

5. Determine which of the following sets of vectors form a basis of the sub-space they span ; find a basis in every case ;

(i) $(1, 3, -3), (2, 5, -2), (1, 1, 6)$ in $V_3(\mathrm{E})$;
(ii) $(3, 1, 2), (4, 1, 3), (-1, -1, 0)$ in $V_3(\mathrm{E})$;
(iii) $(-1, -1, 1), (3, 2, 2), (7, 3, 9)$ in $V_3(\mathrm{R})$;
(iv) $(1, -2, -1, 3), (2, -1, 1, 6), (0, -1, -3, 1)$ in $V_4(\mathrm{R})$.

6. Show that each of the following sets forms a basis of $V_3(\mathrm{R})$:

(i) $(1, -2, 3), (2, 3, 1), (-1, 3, 2)$;
(ii) $(0, 1, 8), (1, -3, 6), (1, -1, 3)$;
(iii) $(0, 3, 2) (2, 0, 1), (1, 5, 4)$.

7. Determine whether the sub-space of $V_4(\mathrm{R})$ spanned by the vectors ϕ_1, ϕ_2, ϕ_3 coincides with the one spanned by the vectors in exercise 5(iv), where

(i) $\phi_1=(3, -1, 2, 9), \phi_2=(1, 3, 6, 2), \phi_3=(1, 2, 5, 2)$;
(ii) $\phi_1=(1, 0, 1, -2), \phi_2=(0, 1, -1, 1), \phi_3=(1, 1, 0, 4)$.

8. Find for $V_4(\mathrm{R})$

(i) a basis which contains the vector $(1, 2, 1, 1)$,
(ii) a basis which contains the vectors $(1, 3, 0, 2)$, and $(1, -1, 2, 0)$,
(iii) a basis containing the vectors $(1, 1, 0, 0), (0, 0, -5, 2)$ and $(0, 2, 3, 0)$,

9. Find the ranks of the following :

(i) Π, defined in exercise 2 (ii), page 32 ;
(ii) the different sub- spaces in exercises 5 (i)-(viii), page 32, and the intersection and the linear sum of every pair of these sub-spaces, taking $n=5$.

10. For $n=5$, find a complement in $V_5(\mathrm{E})$ of each of the different sub-spaces in exercises 5(i)-(viii), page 32.

11. Find in $V_5(\mathrm{R})$ a complement of the sub-space spanned by $(2, -3, 7, 1, 5)$ and $(1, 2, -3, 6, 7)$.

12. If P be a sub-space of a linear space V, over a skew field S, then the factor group $D=V/P$ also forms a linear space over S, if multiplication of the cosets (of P in V) by scalars be defined by

$$c[P+\alpha]=P+c\alpha, \quad \text{where } c \varepsilon S \text{ and } \alpha \varepsilon V;$$

this linear space is called the *factor space* V/P. Show that

(i) if Q be a complement of P in V, and if $\alpha_1, \ldots, \alpha_k$ form a basis of Q, then the cosets $\alpha_1+P, \ldots, \alpha_k+P$ form a basis of the factor space V/P ; and

(ii) $d(V/P) = d(V) - d(P)$.

13. Find the equations of the transformation of co ordinates by a change of basis

(i) from α, β, γ to $\alpha+\beta, \beta+\gamma, \gamma+\alpha$;

(ii) from each basis in exercise 6 to every other one.

3 Normed linear spaces

Let V be a linear space over the field of complex numbers C. V is said to be *normed*, if a real-valued function $\|\alpha\|$, called the *norm* or *length* of α, be defined for every vector $\alpha \varepsilon V$, such that for any $\alpha, \beta \varepsilon V$ and $a \varepsilon C$,

N(1) : $\|\alpha\|>0$, if $\alpha \neq \theta$, (*strictly positive*)

N(2) : $\|a\alpha\| = |a|\|\alpha\|$, and (*homogeneous*)

N(3) : $\|\alpha+\beta\| \leqslant \|\alpha\|+\|\beta\|$. (*sub-additive*)

Now it follows readily that

II.3.1. (*i*) $\|\theta\| = \|0\alpha\| = |0|\|\alpha\| = 0.$

(*ii*) $\|-\alpha\| = \|(-1)\alpha\| = |(-1)|\|\alpha\| = \|\alpha\|.$

Let $\|\alpha-\beta\|$ be called the *distance* between α and β, and denoted by $d(\alpha, \beta)$. Then it follows immediately from $N(1)$, $N(2)$, $N(3)$ and II.3 1 that

II.3.2. *For every pair of vectors* $\alpha, \beta \in V$, *there exists a real number* $d(\alpha, \beta)$, *such that*

(*i*) $d(\alpha, \beta) > 0$, *when* $\alpha \neq \beta$, (*positiveness*)

(*ii*) $d(\alpha, \beta) = d(\beta, \alpha)$, (*symmetry*)

(*iii*) $d(\alpha, \beta) + d(\beta, \gamma) \geqslant d(\alpha, \gamma)$, (*triangle inequality*)

(*iv*) $d(a\alpha, a\beta) = |a| . d(\alpha, \beta)$.

Now, it follows from II.3.2.(i, ii, iii) that

II.3.3. *A normed linear space is metrisable.* †

'79 **Unitary spaces.** Let U be a linear space over the field of complex numbers C. U is said to be a *unitary space, or an inner-product space* if to every pair of vectors α, β, belonging to U, there corresponds a scalar (i.e., a complex number), called the *inner product* or *scalar product* of α, β, and denoted by $\alpha . \beta$, possessing the following properties $SP(1)$, $SP(2)$ and $SP(3)$:

SP(1) : $(\alpha . \beta)^* = \beta . \alpha$, *where* a^* *denotes the conjugate of the complex number* a. (*Hermitian symmetry*)

As an immediate consequence of SP (1) it follows that

II.3.4. *For any vector* $\alpha \in U$, $\alpha . \alpha$ *is a real number.*

Proof : $\alpha . \alpha = (\alpha . \alpha)^*$ is self-conjugate ; hence it is a real number.

This property is further restricted by the assumption ;

SP (2) : $\alpha . \alpha > 0$, *if* $\alpha \neq \theta$. (*positiveness*)

The third property, is given as :

SP (3) : $(a\alpha + b\beta) . \gamma = a(\alpha . \gamma) + b(\beta . \gamma)$, *for any three vectors* α, β, γ *and artitrary scalars* a, b. (*linearity*)

Now one obtains readily the following propositions :

---

† A set M is said to be a *metric space* if for every pair of elements a, b belonging to M, a real number d(a, b) can be defined so as to satisfy the following

M(1) : $d(a, a) = 0$: and $d(a, b) > 0$, if $a \neq b$; (*positiveness*)

M(2) ; $d(a, b) = d(b, a)$: (*symmetry*)

M(3) : $d(a, b) + d(b, c) \geqslant d(a, c)$ (*triangle inequality*).

The real valued function d(a, b) is said to be the *distance function* or the *metric* of the space.

A set in which a suitable metric, satisfying the conditions M(1), M(2), M(3), can be defined is said to be *metrisable*.

II.3.5. *$\alpha.\alpha=0$ if, and only if, $\alpha=\theta$.*

Proof: Putting $\alpha=\theta$, $\gamma=\theta$ and $b=0$ in $SP(3)$, one gets

$$(a\theta+0\beta).\theta=a(\theta.\theta)+0(\beta.\theta)$$

i.e., $$\theta.\theta=a(\theta.\theta), \text{ for every scalar } a.$$

Hence $\theta.\theta=0$. The converse follows from $SP(2)$.

II.3.6. *If $\alpha, \beta, \gamma \in U$ and $a, b \in C$, then*

(i) $(a\alpha).\gamma=a(\alpha.\gamma)$, $\alpha.(b\beta)=b^*(\alpha.\beta)$

(ii) $(\alpha+\beta).\gamma=(\alpha.\gamma)+(\beta.\gamma)$, $\alpha.(\beta+\gamma)=(\alpha.\beta)+(\alpha.\gamma)$.

Proof: (i) The first part follows from SP (3) by taking $b=0$. The second part is obtained as follows.

$$\alpha\,(b\beta,=[(b\beta).\alpha]^*=[b(\beta.\alpha)]^* \quad \text{(from the first part)}$$
$$=b^*(\beta.\alpha)^*=b^*(\alpha.\beta).$$

(ii) The first part follow by taking $a=b=1$ in $SP(3)$. The second part is obtained as follows :

$$\alpha.(\beta+\gamma)=[(\beta+\gamma).\alpha]^*=[(\beta.\alpha)+(\gamma.\alpha)]^* \quad \text{(from the first part)}$$
$$=(\beta.\alpha)^*+(\gamma.\alpha)^*=(\alpha.\beta)+(\alpha.\gamma)$$

The proposition shows that the scalar product of two vectors is *conjugate bilinear* in the two vectors.

By $SP(2)$ and II.3.5, $\sqrt{(\alpha.\alpha)}$ is real. Let the positive square root $\sqrt{(\alpha\,\alpha)}$ be taken as $\|\alpha\|$, the *norm or length* of α, then the conditions $N(1)$, $N(2)$ and $N(3)$ can be verified as follows :

II.3.7. (i) *$\|\alpha\|=0$ if, and only if, $\alpha=\theta$,*

(ii) $\|a\alpha\|=|a|.\|\alpha\|$

Proof: (i) follows immediately from II.3.5

(ii) follows from the fact that

$$\|a\alpha\|=\sqrt{[(a\alpha).(a\alpha)]}=\sqrt{[aa^*(\alpha.\alpha)]}$$
$$=\sqrt{(aa^*)}\sqrt{(\alpha.\alpha)}=|a|\,\|\alpha\|. \quad \text{(by II 3.6 (i))}$$

178

II.3.8 *If $\alpha, \beta \in U$, then $|\alpha.\beta|\leqslant\|\alpha\|\,\|\beta\|$, where the equality holds if, and only if, α and β are linearly dependent.*

(Cauchy-Schwarz inequality).

Proof: If one of the given vectors be null, then the proposition evidently holds – the expressions are equal in this case, and α, β are also linearly dependent. Let α, β be non-null, and let $\beta.\beta = a$ and $\alpha.\beta = b$, then a is real, so that $a = a^*$. Now

$$\begin{aligned}0 &\leqslant (a\alpha - b\beta).(a\alpha - b\beta)\\ &= (a\alpha).(a\alpha) - (a\alpha).(b\beta) - (b\beta).(a\alpha) + (b\beta).(b\beta)\\ &= \|a\alpha\|^2 - [ab^*(\alpha.\beta) + ba(\beta.\alpha)] + \|b\beta\|^2\\ &= a^2\|\alpha\|^2 - (ab^*b + bab^*) + |b|^2\|\beta\|^2\\ &= a^2\|\alpha\|^2 - 2a|b|^2 + |b|^2 a = a[a\|\alpha\| - |b|^2]\\ &= a[\|\alpha\|^2\|\beta\|^2 - |\alpha.\beta|^2].\end{aligned}$$

Dividing both sides of the above inequality by the positive (real) number a, one gets

$$\|\alpha\|^2.\|\beta\|^2 - |\alpha.\beta|^2 \geqslant 0, \text{ i.e., } \|\alpha\|\|\beta\| \geqslant |\alpha.\beta|.$$

Equality holds if, and only if, $a\alpha - b\beta = \theta$, i.e., if, and only if, α and β are linearly dependent.

II.3.9. $\|\alpha + \beta\| \leqslant \|\alpha\| + \|\beta\|$.

Proof:

$$\begin{aligned}\|\alpha+\beta\|^2 &= (\alpha+\beta).(\alpha+\beta) = \alpha.\alpha + \alpha.\beta + \beta.\alpha + \beta.\beta\\ &= \|\alpha\|^2 + [\alpha.\beta + (\alpha.\beta)^*] + \|\beta\|^2\\ &\leqslant \|\alpha\|^2 + 2|\alpha.\beta| + \|\beta\|^2, \quad [\because\ b + b^* \leqslant 2\sqrt{(bb^*)} = 2|b|]\\ &\leqslant \|\alpha\|^2\| + 2\|\alpha\|\|\beta\| + \|\beta\|^2 \quad \text{(by II.3.8.)}\\ &= (\|\alpha\| + \|\beta\|)^2.\end{aligned}$$

It then follows from II.3.7 and II 3 9, that

II.3.10. *A unitary space is normed.*

But every normed linear space is not necessarily unitary. In fact,

II.3.11. Neumann's theorm. *A normed linear space V is unitary if, and only if, for every pair of vectors α, β,*

$$\|\alpha+\beta\|^2 + \|\alpha-\beta\|^2 = 2\|\alpha\|^2 + 2\|\beta\|^2. \qquad \ldots(1)$$

Proof: If V be unitary, then it is normed by (II.3.10) and also

$$\begin{aligned}\|\alpha+\beta\|^2 + \|\alpha-\beta\|^2 &= (\alpha+\beta).(\alpha+\beta) + (\alpha-\beta).(\alpha-\beta)\\ &= 2\alpha.\alpha + 2\beta.\beta = 2\|\alpha\|^2 + 2\|\beta\|^2.\end{aligned}$$

Conversely, let V be a normed linear space, for which the given condition is satisfied. Now, in order that inner product $\alpha.\beta$, satisfying the conditions SP(1-3), may be defined for every pair of vectors α, β in V, such that $\|\alpha\|^2 = \alpha.\alpha$, one must have

$$\|\alpha+\beta\|^2-\|\alpha-\beta\|^2=(\alpha+\beta).(\alpha+\beta)-(\alpha-\beta).(\alpha-\beta)$$
$$=2[\alpha.\beta+\beta.\alpha]=2[\alpha.\beta+(\alpha.\beta)^*]=4R(\alpha.\beta).$$

where $R(\alpha.\beta)$ denotes the real part of $\alpha.\beta$. Then the imaginary part $iI(\alpha.\beta)$ is given by

$$I(\alpha\,\beta)=-R[i(\alpha\,\beta)]=-R[(i\alpha)\,\beta].$$

Thus, finally, the inner product is determined, in terms of the norm function, by

$$\alpha.\beta=R(\alpha.\beta)+iI(\alpha.\beta)=R(\alpha.\beta)-iR[(i\alpha).\beta]$$
$$=[\{\|\alpha+\beta\|^2-\|\alpha-\beta^2\|\}-i\{\|i\alpha+\beta\|^2-\|i\alpha-\beta\|^2\}]/4. \quad \ldots (2)$$

That the inner product, as defined above, satisfies the requisite conditions $SP(1)-SP(3)$ for a unitary space, may now be verified as follows :

(1) $(\alpha.\beta)^*=R(\alpha.\beta)-i.I(\alpha.\beta)=R(\alpha.\beta)+iR[(i\alpha).\beta]$

$$=R(\beta.\alpha)-iR[(i\beta).\alpha]=\beta.\alpha,$$

since $4R(\beta.\alpha)=\|\beta+\alpha\|^2-\|\beta-\alpha\|^2=\|\alpha+\beta\|^2-\|\alpha-\beta\|^2$

$$=4R(\alpha.\beta).$$

and $4R[(i\beta).\alpha]=\|i\beta+\alpha\|^2-\|i\beta-\alpha\|^2$

$$=|-i|^2\|i\alpha-\beta\|^2-|i|^2\|i\alpha+\beta\|^2$$
$$=(-1)[\|i\alpha+\beta\|^2+\|i\alpha-\beta\|^2]=-4R[(i\alpha).\beta].$$

(2) $\alpha.\alpha=R(\alpha.\alpha)-iR[(i\alpha).\alpha]=\|\alpha\|^2>0$, if $\alpha\neq\theta$,

since $R[(i\alpha).\alpha]=\|i\alpha-\alpha\|^2-\|i\alpha+\alpha\|^2$

$$=[|i+1|^2-|i-1|^2]\|\alpha^2=0.$$

(3) The condition $SP(3)$ will be proved in two parts,

$(\alpha+\beta).\gamma=(\alpha.\gamma)+(\beta.\gamma)$ and $(a\alpha).\gamma=a(\alpha.\gamma)$

Now, for any vectors λ, μ, γ,

$$4R[(\lambda+\mu).\gamma]+4R[(\lambda-\mu).\gamma]$$
$$=\|\lambda+\mu+\gamma\|^2-\|\lambda+\mu-\gamma\|^2+\|\lambda-\mu+\gamma\|^2-\|\lambda-\mu-\gamma\|^2$$
$$=[\|(\lambda+\gamma)+\mu\|^2+\|(\lambda+\gamma)-\mu\|^2]-[\|(\lambda-\gamma)+\mu\|^2-\|(\lambda-\gamma)-\mu\|^2]$$
$$=[2\|\lambda+\gamma\|^2+2\|\mu\|^2]-[2\|\lambda-\gamma\|^2+2\|\mu\|^2]$$
$$=2\|\lambda+\gamma\|^2-2\|\lambda-\gamma\|^2=8R(\lambda.\gamma).$$

Hence $R[(\lambda+\mu).\gamma]+R[(\lambda-\mu).\gamma]=2R(\lambda.\gamma).$ ... (3)

Putting $\mu=\lambda$ in (3), one gets $R(2\lambda\,\gamma)=2R(\lambda.\gamma).$ ... (4)

Again putting $\lambda=(\alpha+\beta)/2$ and $\mu=(\alpha-\beta)/2$ in (3),

$$R(\alpha.\gamma)+R(\beta.\gamma)=2R[\{(\alpha+\beta)/2\}.\gamma]$$
$$=R[(\alpha+\beta).\gamma], \text{ by } (4).$$

Now,
$$\begin{aligned}\alpha.\gamma+\beta.\gamma &= R(\alpha.\gamma)-iR(i\alpha.\gamma)+R(\beta.\gamma)-iR(i\beta.\gamma)\\ &= [R(\alpha.\gamma)+R(\beta.\gamma)]-i[R(i\alpha.\gamma)+R(i\beta.\gamma)]\\ &= R[(\alpha+\beta).\gamma]-iR[i(\alpha+\beta).\gamma]\\ &= (\alpha+\beta).\gamma.\end{aligned}$$

To prove the second part, let S be the set of all those complex numbers a for which $(a\alpha).\gamma=a(\alpha.\gamma)$ holds, for any vectors α, γ. Then it follows from the above results that $a, b \varepsilon S$ implies $a+b, a-b \varepsilon S$. Obviously $1 \varepsilon S$; hence all integers, $0, \pm 1, \pm 2, \ldots$, belong to S. Also $a, b \varepsilon S$ and $b \neq 0$ imply $a/b \varepsilon S$ hence all rational numbers belong to S.

Again from $N(3)$, it follows that
$$\|a\alpha+\gamma\|-\|b\alpha+\gamma\| \leqslant \|(a-b)\alpha\|$$
and
$$\|b\alpha+\gamma\|-\|a\alpha+\gamma\| \leqslant \|(b-a)\alpha\|=\|(a-b)\alpha\|.$$
$$\big|\|a\alpha+\gamma\|-\|b\alpha+\gamma\|\big| \leqslant \|(a-b)\alpha\| \to 0,$$
as $a\to b$ (by $N(2)$). Hence $\|a\alpha+\gamma\|$ is a continuous function of a. Likewise, $\|a\alpha-\gamma\|$ is a continuous function of a; hence $R(a\alpha.\gamma)=[\|a\alpha+\gamma\|-\|a\alpha-\gamma\|]/4$, and consequently $a\alpha.\gamma=R(a\alpha.\gamma)-iR(ia\alpha.\gamma)$ are also continuous functions of a. Since every real number can be obtained as the limit of a sequence of rational numbers (by Cantor's theory), it follows that all real numbers belong to S. Also $(i\alpha).\gamma=i(\alpha.\gamma)$, hence $i \varepsilon S$. Then if u, v are real, $u-iv=u+v/i \varepsilon S$. Thus all complex numbers belong to S, so that $(a\alpha).\gamma=a(\alpha.\gamma)$ holds for every complex number a.

Orthogonality of vectors. A vector α is said to be *orthogonal* to a vector β, if $\alpha.\beta=0$ The null vector θ is, by II.3.5, the only vector which is orthogonal to itself. Since $\beta.\alpha=(\alpha.\beta)^*$, orthogonality is a symmetric relation. Also $(a\alpha).(b\beta)=ab^*(\alpha.\beta)$ implies that if α be orthogonal to β, then $a\alpha$ is orthogonal to $b\beta$. Furthermore, if $\beta_k.\alpha=0$, for $k=1, 2, \ldots, m$, then for any scalars $c_1, \ldots, c_m$,
$$(c_1\beta_1+\ldots+c_m\beta_m).\alpha=c_1(\beta_1.\alpha)+\ldots+c_m(\beta_m.\alpha)=0.$$
Thus one obtains :

II.3.12. *If a vector be orthogonal to a given set of vector in a unitary space, then it is orthogonal to every vector belonging to the sub-space spanned by the given set of vectors.*

An immediate consequence of the above proposition is :

II.3.13. *A necessary and sufficient condition that a given vector be orthogonal to every vector of a sub-space is that it should be orthogonal to any set of generators of the sub-space.*

A set of vectors $\gamma_1, \dots, \gamma_m$, is said to be *orthogonal* if

$$\left.\begin{aligned} \gamma_i.\gamma_k &= 0 \text{ for } i \neq k \\ &\neq 0, \text{ for } i = k \end{aligned}\right\} \quad i, k = 1, \dots, m.$$

Thus the vectors belonging to an orthogonal set are pairwise orthogonal non-null vectors.

178 II.3.14. *An orthogonal set of vectors $\gamma_1, \dots, \gamma_m$ is always linearly independent.*

Proof: If $c_1\gamma_1 + \dots + c_m\gamma_m = \theta$, then taking inner product with γ_k,

$$(c_1\gamma_1 + \dots + c_m\gamma_m).\gamma_k = \theta.\gamma_k = 0, \quad i.e., \; c_k\|\gamma_k\|^2 = 0,$$

so that $c_k = 0$, for $k = 1, \dots, m$ Hence the set of vectors, $\gamma_1, \dots, \gamma_k$, is linearly independent.

Orthonormal basis. A vector, whose norm is 1, is said to be a *unit vector* or a *normalised vector*. For every non-null vector α, there exists a corresponding normalised vector $\alpha/\|\alpha\|$.

An orthogonal set of unit vectors is called an *orthonormal set* of vectors. Thus, for an orthonormal set of vectors, $\alpha_1, \dots, \alpha_r$, $\alpha_i.\alpha_j = \delta_{ij}$, where $\delta_{ij} = 0$, for $i \neq j$, and $\delta_{ii} = 1$, for $i, j = 1, \dots, r$.

An orthonormal set of vectors generating a sub-space T is called an *orthonormal basis* of T.

II.3.15. *Any orthonormal set of vectors, $\gamma_1, \dots, \gamma_m$, belonging to a subspace T, can always be extended to an orthonormal basis of T.*

Proof: Let Q be the subspace spanned by the given set of orthonormal vectors $\gamma_1, \dots, \gamma_m$; then Q is a sub-space of T. If $Q = T$, then the conclusion is obvious Otherwise, there exists a non-null vector $\delta \in T - Q$ Then the vector

$$\lambda = \delta - c_1\gamma_1 - \dots - c_m\gamma_m$$

is orthogonal to each of the vectors $\gamma_1, \dots, \gamma_m$, if

$$0 = \lambda.\gamma_k = \delta.\gamma_k - (c_1\gamma_1 + \dots + c_m\gamma_m).\gamma_k = \delta.\gamma_k - c_k.$$

i.e., if $c_k = \delta.\gamma_k$, for $k = 1, \dots, m$.

The vector $\lambda \neq \theta$, otherwise $\delta = c_1\gamma_1 + \ldots + c_m\gamma_m \varepsilon Q$. Let $\gamma_{m+1} = \lambda/\|\lambda\|$, then $\gamma_1, \ldots, \gamma_m, \gamma_{m+1}$ form an orthonormal set, and they generate a sub-space R of rank $m+1 = d(Q)+1$, where $Q \subset R \subset T$ (each being a sub-space of the next one).

The proof of the proposition is now completed by repeating the process $d(T) - d(Q)$ times.

This process of extending an orthonormal set is known as *Gramm-Schmidt orthogonalisation process.*

For a given sub-space T (different from the null-space), one can always find a unit vector by normalising any non-null vector in T; and then by applying the above process, one obtains an orthonormal basis of T; thus

II.3.16. *Any (non-null) sub-space of a unitary space possesses an orthonormal basis.*

Orthogonal complement. Two subsets A and B of a unitary space E are said to be *orthogonal*, if each vector of A is orthogonal to every vector in B. Since θ is the only vector in E, which is orthogonal to itself, it follows that $A \cap B$ is either empty or consists of θ only.

II.3.17. *If P be a sub-space of a unitary space E, and if Q be the set of all those vectors in E which are orthogonal to every vector in P, then Q is a sub-space of E, and E is the direct sum of P and Q.*

Proof: Let $\alpha_1, \ldots, \alpha_m$ be an orthonormal basis of P. The orthonormal set, $\alpha_1, \ldots, \alpha_m$, may then be extended to an orthonormal basis, $\alpha_1, \ldots, \alpha_m, \alpha_{m+1}, \ldots, \alpha_n$ of the space E. Let Q be the sub-space spanned by the vectors $\alpha_{m+1}, \ldots, \alpha_n$, then every vector in Q is orthogonal to every vector in P, since

$$(a_1\alpha_1 + \ldots + a_m\alpha_m).(b_{m+1}\alpha_{m+1} + \ldots + b_n\alpha_n) = 0.$$

Conversely, if a vector $\xi = c_1\alpha_1 + \ldots + c_n\alpha_n$ in E be orthogonal to every vector in P, then $\xi.\alpha_k = 0$, i.e., $c_k = 0$, for $k = 1, \ldots, m$; thus $\xi = c_{m+1}\alpha_{m+1} + \ldots + c_n\alpha_n \varepsilon Q$. Hence Q consists of all those vectors in E which are orthogonal to every vector in P.

Also any vector $\xi \varepsilon E$ can be expressed uniquely as

$$\xi=(c_1\alpha_1+\ldots+c_m\alpha_m)+(c_{m+1}\alpha_{m+1}+\ldots+c_n\alpha_n)$$
$$=\sigma+\tau, \quad \text{where } \sigma \varepsilon P \text{ and } \tau \varepsilon Q;$$

hence E is the linear sum of P and Q, i.e., $E=P+Q$.

The null vector θ being the only vector, which is orthogonal to itself, it follows that $P\cap Q=\theta$.

Consequently E is the direct sum of P and Q.

The sub-space Q is called the *orthogonal complement of* P *in* E ; Q is *uniquely* determined. Thus, whereas a sub-space P has many complements in E, the orthogonal complement is unique.

It now follows from II.3.17 that

II.3.18. *A unitary space* E *is the direct sum of any linear sub-space and its orthogonal complement in* E.

It has been seen that every vector $\xi \varepsilon E$ has a *unique decomposition* $\xi=\sigma+\tau$, where $\sigma \varepsilon P$ and $\tau \varepsilon Q$; σ is called the *orthogonal projection of* ξ *on* P, *and* τ is called the *component of* ξ *orthogonal to* P.

If $A_1,\ldots, A_m$ be a set of pairwise orthogonal sub-spaces of E, then their linear sum $A_1+\ldots+A_m$ is called their *orthogonal sum*. It then follows immediately that :

II.3.19. *If* T *be the orthogonal sum of the sub-spaces* $A_1,\ldots, A_m$ *in a unitary space* E, *then*

(*i*) *an orthonormal basis of* A_1, *an orthonormal basis of* $A_2,\ldots$, *and an orthonormal basis of* A_m, *taken together, constitute an orthonormal basis of* T ;

(*ii*) *every vector* $\xi \varepsilon T$ *can be expressed uniquely as a sum*, $\xi=\alpha_1+\ldots+\alpha_m$, *where* $\alpha_1 \varepsilon A_i$, *for* $i=1,\ldots, m$;

(*iii*) T *is the direct sum of* $A_1,\ldots, A_m$, *and*

$$d(T)=d(A_1)+\ldots+a(A_m).$$

Co-ordinatisation and transformation of co-ordinates.

II.3.20. *Let* $(x_1,\ldots, x_n)$ *and* $(y_1,\ldots, y_r)$ *be the co-ordinate vectors corresponding to vectors* ξ,η *in a unitary space* E *with* *respect to an orthonormal basis, then their scalar product is given by* $\xi.\eta = x_1 y_1^* + \dots + x_n y_n^*$.

Proof: Let $\alpha_1,\dots,\alpha_n$ be the given orthonormal basis of E, then

$$\xi = x_1\alpha_1 + \dots + x_n\alpha_n,\quad \eta = y_1\alpha_1 + \dots + y_n\alpha_n.$$

Hence
$$\begin{aligned}\xi.\eta &= (x_1\alpha_1 + \dots + x_n\alpha_n).(y_1\alpha_1 + \dots + y_n\alpha_n)\\ &= x_1 y_1^* + \dots + x_n y_n^*,\end{aligned}$$

since $\alpha_i.\alpha_k = 0$, when $i \neq k$, and $\alpha_k.\alpha_k = 1$, for $i, k = 1,\dots,n$.

It now follows that, when referred to an orthonormal basis, the length of a vector $\xi = (x_1,\dots, x_n)$ is given by

$$\|\xi\|^2 = \xi.\xi = x_1 x_1^* + \dots + x_n x_n^* = |x_1|^2 + \dots + |x_n|^2.$$

There exist different orthonormal bases for the same subspace. For a change from one orthonormal basis to another, one obtains the following from II.2.10.

II.3.21. *Let* $x_k' = \sum_{i=1}^{n} x_i a_{ik}$, $k = 1,\dots, n$, *be the equations of the transformation of co-ordinate vectors induced by a change from an orthonormal basis* $\alpha_1,\dots,\alpha_n$ *to an orthonormal basis* $\beta_1,\dots,\beta_n$, *then* $a_{i1}a_{m1}^* + \dots + a_{in}a_{mn}^* = \delta_{im}$, *where* $\delta_{im} = 0$ *for* $i \neq m$, *and* $\delta_{ii} = 1$, *for* $i, m = 1,\dots,n$.

Proof: If follows from the result (2), page 38, that

$$\alpha_i = a_{i1}\beta_1 + \dots + a_{in}\beta_n, \quad \text{for } i = 1,\dots,n.$$

Thus
$$\begin{aligned}\alpha_i.\alpha_m &= (a_{i1}\beta_1 + \dots + a_{in}\beta_n).(a_{m1}\beta_1 + \dots + a_{mn}\beta_n)\\ &= a_{i1}a_{m1}^* + \dots + a_{in}a_{mn}^*. \qquad [\because\ \beta_i\beta_k = \delta_{ik}]\end{aligned}$$

But $\alpha_i \alpha_m = \delta_{im}$; hence the result.

Exercises

1. In a normed linear space V, over the field of complex numbers, let

$$N(\alpha, \beta) = [\|\alpha+\beta\|^2 + \|\alpha-\beta\|^2]/2[\|\alpha\|^2 + \|\beta\|^2],$$

and if λ be the greatest lower bound of $N(\alpha, \beta)$ and μ be the least upper bound of $N(\alpha, \beta)$ for all $\alpha, \beta \in V$, then

$$\tfrac{1}{2} \leqslant \lambda \leqslant 1 \leqslant \mu \leqslant 2 \text{ and } \lambda\mu = 1.$$

Thus V is unitary if, and only if, $\lambda = \mu = 1$.

2. In a unitary space U, a vector α is the null vector if, and only if, $\alpha.\xi=0$ for every $\xi \in U$.

3. If $\alpha_1, \alpha_2, \ldots, \alpha_m$ be a set of vectors in a unitary space, then the following determinant, known as the *Grammian*,

$$\begin{vmatrix} \alpha_1.\alpha_1 & \alpha_1.\alpha_2 & \ldots & \alpha_1.\alpha_m \\ \alpha_2.\alpha_1 & \alpha_2.\alpha_2 & \ldots & \alpha_2.\alpha_m \\ \ldots & \ldots & \ldots & \ldots \\ \alpha_m.\alpha_1 & \alpha_m.\alpha_2 & \ldots & \alpha_m.\alpha_m \end{vmatrix} \geqslant 0,$$

where the value is 0 if, and only if, the given set is linearly dependent.

4. If α be orthogonal to β, in a unitary space E, then

$$\|\alpha+\beta\|^2=\|\alpha\|^2+\|\beta\|^2. \qquad \textit{(Pythagorean property)}$$

5. If $\alpha_1, \ldots, \alpha_n$ be an orthonormal set of vectors in a unitary space E, then

(i) $(\xi.\alpha_1)^2+\ldots+(\xi.\alpha_n)^2 \leqslant \|\xi\|^2$, *(Bessel's inequality)*

(ii) $(\xi.\alpha_1)(\eta.\alpha_1)+\ldots+(\xi.\alpha_n)(\eta.\alpha_n)=\xi.\eta$, *(Perseval's identity)*

where $\xi, \eta \in E$.

6. Show that all complex valued functions, continuous on the real interval $0 \leqslant t \leqslant 1$, form an infinite dimensional unitary space, when the inner product of two functions f and g is defined by

$$f.g=\int_0^1 f(t)[g(t)]^* dt,$$

where * indicates the conjugate.

7. **Quasi-normed space.** Let V be a linear space over the field of real numbers **E**, such that a real-valued function (*functional*) $\|\alpha\|$, called the *quasi-norm* of α, is defined for every vector $\alpha \in V$, satisfying the following conditions :

Q(1) : $\|\alpha\| \geqslant 0$, for every $\alpha \in V$,
Q(2) : $|a| \leqslant |b|$ implies $\|a\alpha\| \leqslant \|b\alpha\|$,
Q(3) : $\|\alpha+\beta\| \leqslant \|\alpha\| + \|\beta\|$,
Q(4) : $\|p\alpha\|$ tends to 0, as p tends to 0,

where $\alpha, \beta \in V$ and $a, b, p \in$ **E** ; then V is called a *quasi-normed space.*

Quasi-metric space. Putting $\|\alpha-\beta\|=q(\alpha, \beta)$, one obtains a *quasi-metric* (or *ecart* or *pseudo-metric*) in V. Show that V froms a *quasi-metric space*, in which

(i) $q(\alpha, \alpha)=0$, $q(\alpha, \beta)\geqslant 0$,

(ii) $q(\alpha, \beta)=q(\beta, \alpha)$,

(iii) $q(\alpha, \beta)+q(\beta, \gamma)\geqslant q(\alpha, \gamma)$.

A quasi-metric space becomes a metric space, when $q(\alpha,\beta)=0$ implies $\alpha=\beta$.

N.B. The terms *ecart* and *pseudo-metric* are also used in a slightly different sense.

8. **Pseudo-normed space.** A *pseudo-norm* in V (over E) is determined by the following properties :

$Ps(1)$: $\|\alpha\|\geqslant 0$ for every $\alpha \in V$,

$Ps(2)$: $\|a\alpha\|=|a|\,\|\alpha\|$ for every real number a.

V is then called a *pseudo-normed space*.

A pseudo-norm is a norm if it satisfies the two conditions :

(1) $\|\alpha\|=0$ implies $\alpha=\theta$, (2) $\|\alpha+\beta\|\leqslant\|\alpha\|+\|\beta\|$.

3.1. Euclidean linear spaces.

A finite dimensional unitary space V, over the field of real numbers, is called a *Euclidean linear space*. Thus the scalar product $\alpha.\beta$ for every pair of vectors $\alpha, \beta \in V$ is a real number, and it is symmetric with respect to α, β, i.e., $\beta.\alpha=\alpha.\beta$, All the properties, proved for a unitary space, hold also for a Euclidean linear space (where $a^*=a$ for every scalar a involved in those properties).

The concepts of length, distance and orthogonality have already been introduced in a unitary space. The magnitude of an angle may now be defined more concretely in a Euclidean linear space. From Cauchy-Schwarz inequality, it follows that, for any pair of non-null vectors $\alpha, \beta \in V$, $-1\leqslant(\alpha\beta)/(\|\alpha\|\,\|\beta\|)\leqslant 1$. Hence,

$$(\alpha.\beta)/(\|\alpha\|\,\|\beta\|)=\cos\phi, \qquad \text{for } 0\leqslant\phi\leqslant\pi,$$

where ϕ is uniquely determined ; ϕ is defined to be the magnitude of the *angle between the vectors* α and β. Thus $\phi=\pi/2$ if, and only if, $\alpha.\beta=0$, *i.e.*, if α and β are orthogonal.

Exercises

1. Prove the analogues of the propositions II.3.14. II.8.15 and II.3.16 in a Euclidean linear space.

2. Let λ and μ be two given vectors in a Euclidean linear space. Find the value of u for which $\lambda + u\mu$ has the shortest length.

3. In any Euclidean linear space, $\alpha.\beta = 0$ if, and only if,

$$\|\alpha + c\beta\| \geqslant \|\alpha\|,$$

for every real number c.

4. Let P° denote the orthogonal complement of a subspace P in a Euclidean linear space E. Show that

(i) $(P^{\circ})^{\circ} = P$,

(ii) $(P+Q)^{\circ} = P^{\circ} \cap Q^{\circ}$,

(iii) $(P \cap Q)^{\circ} = P^{\circ} + Q^{\circ}$,

where P, Q are any two subspaces of E.

5. Obtain an orthonormal basis of the 4-dimensional Euclidean linear space E^4 having

(*i*) $\alpha = (0, 1/\sqrt{2}, -1/\sqrt{2}, 0)$ and $\beta = (0, 0, 0, 1)$;

(*ii*) $\alpha = (\mathrm{Cos}\ \psi, 0, \mathrm{Sin}\ \psi, 0)$ and

$\beta = (\mathrm{Sin}\ \psi\ \mathrm{Cos}\ \phi, \mathrm{Sin}\ \psi\ \mathrm{Sin}\ \phi, 0, \mathrm{Cos}\ \psi)$

as two of the basis vectors.

If P be the sub-space generated by α, β, obtain in both cases the orthogonal complement of P in E^4.

6. If a Euclidean linear space E be the direct sum of two sub-space P and Q, then an orthonormal basis of P together with an orthonormal basis of Q form an orthonormal basis of E.

7. Find the equations of transformation of co-ordinate vectors in E^3 induced by a change from the basis

$$\left.\begin{array}{l} \alpha_1 = (1, 0, 0), \\ \alpha_2 = (0, 1/\sqrt{2}, -1/\sqrt{2}), \\ \alpha_3 = (0, 1/\sqrt{2}, 1/\sqrt{2}) \end{array}\right\} \text{to the basis} \left\{\begin{array}{l} \beta_1 = (1/\sqrt{3}, -1/\sqrt{3}, 1/\sqrt{3}) \\ \beta_2 = (1/\sqrt{2}, 1/\sqrt{2}, 0), \\ \beta_3 = (-1/\sqrt{6}, 1/\sqrt{6}, 2/\sqrt{6}). \end{array}\right.$$

CHAPTER III

MATRICES

Let R be a ring. A rectangular array of m rows and n columns of elements from R is called an *$m \times n$ matrix over R* :

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}. \qquad \text{...(1)}$$

It is also denoted by $A = (a_{ik})$, a_{ik} being the element in the (i,k)th position (i.e., it belongs to the ith row and at the same time to the kth column). All the elements of the matrix are obtained by taking $i = 1, 2, \ldots, m$ and $k = 1, 2, \ldots, n$.

In particular, the $1 \times n$ matrices $a_i = (a_{i1}, a_{i2}, \ldots, a_{in})$, for $i = 1, 2, \ldots, m$, are called the *row-vectors* of the matrix A, whereas the $m \times 1$ matrices

$$\alpha^{(k)} = \begin{bmatrix} a_{1k} \\ a_{2k} \\ \\ a_{mk} \end{bmatrix}, \qquad \text{for } k = 1, 2, \ldots, n,$$

are the *column-vectors* of A. Having obtained the column-vectors, as above, one can as well write them horizontally as as $\alpha^{(k)} = (a_{1k}, a_{2k}, \ldots, a_{mk})$.

A matrix obtained by omitting some rows or some columns or both from a given matrix A is called a *sub-matrix* or a *block* of A.

1. Operations on matrices.

Let R be a given ring, and $A = (a_{ik})$, $B = (b_{ik}), \ldots$, be matrices over R.

Addition of matrices. Addition of two matrices is defined when the two matrices have the same size (i e., have same number of rows and the same number of columns). Let A and B be two $m \times n$ matrices, then their sum is defined to be an $m \times n$ matrix $S=(s_{ik})$. i.e., $S=A+B$, where

$$s_{ik}=a_{ik}+b_{ik}, \text{ for } i=1,\ldots,m, \text{ and } k=1,\ldots, n.$$

III.1.1. *The set of all $m \times n$ matrices over a ring R forms an abelian group.*

Proof: Let Ω be the set of all $m \times n$ matrices over R, then Ω is closed with respect to addition. The associative property of addition holds in Ω, since

$$[(a_{ik})+(b_{ik})]+(c_{ik})=(a_{ik}+b_{ik})+(c_{ik})=([a_{ik}+b_{ik}]+c_{ik})$$
$$=(a_{ik}+[b_{ik}+c_{ik}]) \qquad \text{(by associative property in } R)$$
$$=(a_{ik})+(b_{ik}+c_{ik})=(a_{ik})+[(b_{ik})+(c_{ik})].$$

The commutative property of addition is satisfied in Ω, since

$$(a_{ik})+(b_{ik})=(a_{ik}+b_{ik})=(b_{ik}+a_{ik})=(b_{ik})+(a_{ik}).$$

The matrix O, all whose elements are the null element of R, is the null element in Ω, $O+(a_{ik})=(a_{ik})$, for all i, k. The negative of a matrix (a_{ik}) is the matrix $(-a_{ik})$. Hence Ω is an abelian group.

The matrix O is called *the null matrix.*

Multiplication of a matrix by a scalar. The product of an $m \times n$ matrix (a_{ik}) by a scalar $c \in R$ is defined to be an $m \times n$ matrix (b_{ik}), where $b_{ik}=c.a_{ik}$, for $i=1,\ldots, m$ and $k=1,\ldots, n$. In particular, $0.A=O$, for every matrix A.

Taking R to be, in particular, a skew field S, one obtains:

III.1 2. *The set of all $m \times n$ matrices over a skew field S forms a linear space of rank mn over S.*

Proof: The set Ω of all $m \times n$ matrices over S is an abelian group, and the product of a matrix by a scalar is also a matrix in Ω. Also

$$1.(a_{ik})=(1.a_{ik})=(a_{ik}),$$
$$c[d(a_{ik})]=c(da_{ik})=(cda_{ik})=cd(a_{ik}),$$
$$c[(a_{ik})+(b_{ik})]=c(a_{ik}+b_{ik})=(c[a_{ik}+b_{ik}])=(ca_{ik}+cb_{ik}).$$
$$=(ca_{ik})+(cb_{ik})=c(a_{ik})+c(b_{ik}).$$

and $[c+d](a_{ik})=([c+d]a_{ik})=(ca_{ik}+da_{ik})=(ca_{ik})+(da_{ik})$
$=c(a_{ik})+d(a_{ik})$.

Hence Ω is a linear space over S.

Again, let O_{ik} be be $m\times n$ matrix, of which the (i,k)th element is 1 and all other elements are 0. Then there are exactly mn such matrices, and any matrix (a_{ik}) can be expressed as a linear combination of these matrices :

$$(a_{ik})=\sum_{i=1}^{m}\sum_{k=1}^{n} a_{ik}O_{ik} \qquad \ldots(1)$$

Hence the matrices O_{ik} form a set of generators of Ω. Also, a matrix (a_{ik}) is the null matrix if, and only if, a_{ik} are all null ; hence it follows from (1) that the matrices O_{ik} are linearly independent over S. Thus, the mn matrices O_{ik} form a basis of Ω ; consequently, Ω has the rank mn.

Multiplication of matrices. Product of two matrices A and B is defined only when the number of columns in A is the same as the number of rows in B ; in that case the pair of matrices A, B is said to be *conformable*. Let A be an $m\times n$ matrix and B an $n\times t$ matrix, then their product, $AB=P=(p_{ik})$, is determined as an $m\times t$ matrix, where

$$p_{ik}=a_{i1}b_{1k}+a_{i2}b_{2k}+\ldots+a_{in}b_{nk}=\sum_{j=1}^{n}a_{ij}b_{jk},$$

for $i=1,\ldots,m$ and $k=1,\ldots,t$. Thus p_{ik} is the scalar product of the vectors α_i and $\beta^{(k)}$, where $\alpha_i=(a_{i1},\ldots,a_{in})$, for $i=1\ldots,m$, are the row vectors of A, and $\beta^{(k)}=(b_{1k},\ldots,b_{nk})$, for $k=1,\ldots,t$, are the column-vectors of B. So, the elements of the product matrix are obtained by the following *row-by-column rule* :

"*The (i,k)th element of the product matrix is obtained by taking the scalar product of the ith row-vector of the left matrix and the kth column-vector of the right matrix.*"

III.1.3. *Multiplication of matrices is associaive and distributive.*

Proof : Let A be an $m\times n$ matrix, B an $n\times t$ matrix, and C a $t\times s$ matrix. Then the product matrices $AB=P$, $(AB)C=Q$,

$BC=G$ and $A(BC)=H$ are defined, and

$$p_{ik}=\sum_{u=1}^{n} a_{iu}c_{uk}, \text{ for } i=1,\ldots,m, \text{ and } k=1,\ldots,t,$$

$$g_{ik}=\sum_{u=1}^{t} b_{iu}b_{uk}, \text{ for } i=1,\ldots, n, \text{ and } k=1,\ldots, s\ ;$$

hence
$$q_{ik}=\sum_{r=1}^{t} p_{ir}c_{rk}=\sum_{r=1}^{t}\Big[\sum_{u=1}^{n} a_{iu}b_{ur}\Big]c_{rk}=\sum_{u=1}^{n} a_{iu}\Big[\sum_{r=1}^{t} b_{ur}c_{rk}\Big]$$

$$=\sum_{u=1}^{n} a_{iu}g_{uk}=h_{ik}, \text{ for } i=1,\ldots,m, \text{ and } k=1,\ldots, s.$$

Thus $(AB)C=A(BC)$, which proves the associative property.

Next, let D be an $m\times n$ matrix, then the matrix products $AB=P$, $DB=V$ and $(A+D)B=W$ are defined, where

$$w_{ik}=\sum_{u=1}^{n}[a_{iu}+d_{iu}]b_{uk}=\sum_{u=1}^{n} a_{iu}b_{uk}+\sum_{u=1}^{n} d_{iu}b_{uk}$$

$$=p_{ik}+v_{ik}, \text{ for } i=1,\ldots,m, \text{ and } k=1,\ldots, t.$$

Thus $(A+D)B=AB+DB$; similarly, for a given $j\times m$ matrix F, one obtains $F(A+D)=FA+FD$.

It should however be noted that matrix multiplication is *not commutative.* In the first place, for a conformable pair A, B, the pair B, A is not conformable, unless these are square matrices. Secondly, even for two square matrices, multiplication is not, in general, commutative ; e g.

$$\begin{pmatrix}2 & -1\\ 3 & 0\end{pmatrix}\begin{pmatrix}1 & 3\\ -4 & 2\end{pmatrix}=\begin{pmatrix}6 & 4\\ 3 & 9\end{pmatrix}, \quad \begin{pmatrix}1 & 3\\ -4 & 2\end{pmatrix}\begin{pmatrix}2 & -1\\ 3 & 0\end{pmatrix}=\begin{pmatrix}11 & -1\\ -2 & 4\end{pmatrix}.$$

Transposition of a matrix. Let $A=(a_{ik})$ be an $m\times n$ matrix, then the $n\times m$ matrix $P=(p_{ik})$, where $p_{ik}=a_{ki}$, for $i=1,\ldots, n$ and $k=1,\ldots, m$, is called the *transpose of the matrix* A ; this is expressed symbolically by the notation $P=A^t$. Thus the transpose matrix A^t is derived from A by changing the rows of A to the corresponding columns of A^t (and consequently, the columns of A become the rows of A^t). The effect of transposition on the composition of matrices is considered in the following :

III.1.4. *(i)* $(A+B)^t = A^t + B^t$, *(ii)* $(AC)^t = C^t A^t$, *and* *(iii)* $(gA)^t = gA^t$, *g being a scalar.*

Proof : *(i)* The ith row-vector of $(A+B)^t$

$=i$th column-vector of $A+B$

$=i$th column-vector of $A+i$th column-vector of B

$=i$th row-vector of A^t+ith row vector B^t

$=i$th row-vector of (A^t+B^t).

(ii) The (i,j)th element of $(AC)^t=(j,i)$th element of AC

$=(j$th row-vector of $A).(i$th column-vector of $C)$

$=(j$th column-vector of $A^t).(i$th row-vector of $C^t)$

$=(i$th row-vector of $C^t)(j$th column-vector of $A^t)$

(since the scalar product is commutative)

$=(i, j)$th element of C^tA^t.

(iii) The ith row-vector of $(gA)^t = i$th column-vector of gA

$=g.(i$th column-vector of $A)=g.(i$th row-vector of $A^t)$

$=i$th row-vector of gA^t.

Block-multiplication. Let A, B be a conformable pair of matrices. The two matrices may be considered to be composed of blocks or submatrices A_{ik}, B_{ik}. The product matrix AB may then be obtained by applying the row-by-column rule on the rows of blocks of A and the columns of blocks of B. Consider for example,

$$A=\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}, \quad B=\begin{bmatrix} B_{11} & B_{12} & B_{13} & B_{14} \\ B_{21} & B_{22} & B_{23} & B_{24} \\ B_{31} & B_{31} & B_{33} & B_{34} \end{bmatrix}$$

then $$AB=\begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \end{bmatrix},$$

where the blocks of AB are obtained as follows :

$C_{ik}=A_{i1}B_{1k}+A_{i2}B_{2k}+A_{i3}B_{3k}$, for $i=1, 2$, and $k=1, 2, 3, 4$.

This presumes of course that the blocks of A and B are of suitable sizes such that the products of the submatrices, in the above expressions, are defined. Hence the pairs of blocks A_{ir}, B_{rk} must be conformable, for $i=1,2$, $k=1, 2, 3, 4$ and $r=1, 2, 3$.

To prove the rule of block multiplication, let $(A_{ik}, 0)$ be the matrix obtained from A by retaining the block A_{ik} and replacing all other elements by the null element ; the matrices $(B_{ik}, 0)$ are defined in a similar way. Then

$$A=(A_{11}, 0)+(A_{12}, 0)+(A_{13}, 0)+(A_{21}, 0)+(A_{22}, 0)+(A_{23}, 0)$$

$$B=(B_{11}, 0)+\ldots+(B_{14}, 0)+(B_{21}, 0)+\ldots+(B_{24}, 0)+(B_{31}, 0)+\ldots+(B_{34}, 0).$$

Hence by the distributive property, if follows that

$$AB=\Big[\sum_{i=1}^{2}\sum_{r=1}^{3}(A_{ir}, 0)\Big]\Big[\sum_{r=1}^{3}\sum_{k=1}^{4}(B_{rk}, 0)\Big]=\sum_{i=1}^{2}\sum_{k=1}^{4}(C_{ik}, 0),$$

where C_{ik} has the expression as given above.

EXAMPLE 1 : To find the product of the following matrices by the rule of block multiplication (considering the blocks as indicated) :

$$A=\left[\begin{array}{cc|c} 2 & 1 & -1 \\ \hline 1 & 0 & 2 \\ 3 & 4 & -3 \end{array}\right], \qquad B=\left[\begin{array}{c|ccc} 1 & -1 & 0 & 2 \\ -3 & 2 & 1 & 3 \\ \hline 2 & -2 & 4 & 1 \end{array}\right].$$

$$AB=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}=\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix},$$

where $C_{11}=A_{11}B_{11}+A_{12}B_{21}=(2, 1)\begin{pmatrix} 1 \\ -3 \end{pmatrix}+(-1)(2)=-1-2=-3,$

$$C_{12}=A_{11}B_{12}+A_{12}B_{22}=(2, 1)\begin{pmatrix} -1 & 0 & 2 \\ 2 & 1 & 3 \end{pmatrix}+(-1)(-2, 4, 1)$$

$$=(0, 1, 7)+(2, -4, -1)=(2, -3, 6).$$

$$C_{21}=A_{21}B_{11}+A_{22}B_{21}=\begin{pmatrix} 1 & 0 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 1 \\ -3 \end{pmatrix}+\begin{pmatrix} 2 \\ -3 \end{pmatrix}(2)$$

$$=\begin{pmatrix} 1 \\ -9 \end{pmatrix}+\begin{pmatrix} 4 \\ -6 \end{pmatrix}=\begin{pmatrix} 5 \\ -15 \end{pmatrix}$$

$$C_{22}=A_{21}B_{12}+A_{22}B_{22}=\begin{pmatrix} 1 & 0 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} -1 & 0 & 2 \\ 2 & 1 & 3 \end{pmatrix}+\begin{pmatrix} 2 \\ -3 \end{pmatrix}(-2, 4, 1)$$

$$=\begin{pmatrix} -1 & 0 & 2 \\ 5 & 4 & 18 \end{pmatrix}+\begin{pmatrix} -4 & 8 & 2 \\ 6 & -12 & -3 \end{pmatrix}=\begin{pmatrix} -5 & 8 & 4 \\ 11 & -8 & 15 \end{pmatrix}.$$

Hence
$$AB = \left[\begin{array}{r|rrr} -3 & 2 & -3 & 6 \\ \hline 5 & -5 & 8 & 4 \\ -15 & 11 & -8 & 15 \end{array}\right].$$

Considering a different partition of the matrices into blocks, one obtains in a similar manner :

$$\left[\begin{array}{r|rr} 2 & 1 & -1 \\ 1 & 0 & 2 \\ \hline 3 & 4 & -3 \end{array}\right] \cdot \left[\begin{array}{rr|rr} 1 & -1 & 0 & 2 \\ \hline -3 & 2 & 1 & 3 \\ 2 & -2 & 4 & 1 \end{array}\right] = \left[\begin{array}{rr|rr} -3 & 2 & -3 & 6 \\ 5 & -5 & 8 & 4 \\ \cline{3-4} -15 & 11 & -8 & 15 \end{array}\right].$$

Exercises

1. For the following matrices,

$$L = \begin{bmatrix} 3 & -2 & 1 \\ 2 & -1 & 1 \\ -2 & 2 & 0 \end{bmatrix}, \quad M = \begin{bmatrix} 1 & -2 \\ 1 & -2 \\ -1 & 2 \end{bmatrix}, \quad N = \begin{pmatrix} 4 & -4 & 2 \\ 2 & -2 & 1 \end{pmatrix}.$$

find the matrices $M+N^t$, M^t+N, LM, MN, NL, LN^t, M^tL, L^2-L, and $(L-I)^2-2L+3I$.

2. Show that $A^3+2A^2-23A-37I=0$, where

$$A = \begin{bmatrix} 2 & 1 & -3 \\ 1 & -5 & 0 \\ -3 & -1 & 1 \end{bmatrix}.$$

3. Show that the three matrices **U**, **V**, **W**, have the properties of the three quaternions i, j, k, where

$$U = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \quad W = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

4. Show that for the spin matrices,

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

(*i*) $X^2=Y^2=Z^2=I$, and

(*ii*) any product of the matrices **I**, **X**, **Y**, **Z** is equal to one of these matrices multiplied by a scalar.

5. Find all the matrices H, such that GH=HG, where

$$(i)\ G=\begin{bmatrix}0&0&1\\0&1&0\\1&0&0\end{bmatrix},\qquad (ii)\ G=\begin{bmatrix}2&0&0\\0&3&3\\0&0&4\end{bmatrix}.$$

2. Equivalence of matrices.

Let the ground system be taken to be, in particular, a skew field K.

The *elementary row-operations* on a matrix A, over K, are defined to be the following :

(I) interchange of any two row-vectors ;

(II) multiplication of a row-vector by a non-null scalar ;

(III) replacement of any row-vector by the sum of that row-vector and c-times any other row-vector i.e., replacement of α_r by $\alpha_r+c\alpha_k$, where c in a non-null scalar.

Row-equivalence. Let A, B be two $m\times n$ matrices, over K, then B is said to be *row-equivalent* to A, if B can be obtained from A by applying some elementary row-operations. Since for every elementary row-operation, there exsits an inverse elementary row-operation (of the same type), it follows that row-equivalence is a symmetric-relation. Row-equivalence being obviously reflexive and transitive, it is an equivalence relation. Hence the aggregate of all $m\times n$ matrices over K is partitioned into disjoint classes of row-equivalent matrices.

III,2.1. *Any matrix A, over a skew field K, is row equivalent to a matrix in the echelon form:*

$$E=\begin{bmatrix}1&*&\cdots&*&*&\cdots&*\\0&1&\cdots&*&*&\cdots&*\\\cdots&\cdots&\cdots&\cdots&&&\cdots\\0&0&\cdots&1&*&\cdots&*\\0&0&\cdots&0&0&\cdots&0\\\cdots&\cdots&\cdots&\cdots&&&\cdots\\0&0&\cdots&0&0&\cdots&0\end{bmatrix},\qquad\ldots(1)$$

or a matrix which differs from the above form by a permutation of the columns only ; the asterisks denote arbitray scalars.

Proof : Let the given matrix A be taken as given in (1), page 55. The null-columns, i.e., columns all whose elements are null, remain so by any elementary row-operations. Such columns will not be considered in the numbering of the columns ; that is, by the first column, the second column, etc. will be meant the first non-null column, the second non-null column, etc. Now, since all the elements in the first column are not null, one can bring a non-null element in the (1, 1)—position by an elementary row-operation of the type I, so that the element a_{11} may be supposed to be non-null. Then applying elementary row-operations of types II and III, the row-vectors $\alpha_1, \alpha_2, \ldots \alpha_m$ are replacee by the vectors

$$\beta_1=a_{11}^{-1}\alpha_1,\ \beta_2=\alpha_2-a_{21}a_{11}^{-1}\alpha_1, \ldots,\ \beta_m=\alpha_m-a_{m1}a_{11}^{-1}\alpha_1.$$

Taking $\beta_k=(b_{k1}, b_{k2}, \ldots, b_{kn})$, for $k=1, \ldots, m$, it is found that $b_{11}=1$ and $b_{21}=b_{31}=\ldots=b_{m1}=0$. Let B be the matrix having $\beta_1, \ldots, \beta_m$ as its row-vectors, then its first column vector is $\beta^{(1)}=(1, 0, \ldots, 0)$. The matrix B is said to have been obtained by *sweeping out* the first column of A, and it is row-equivalent to A.

Let B_1 be the $(m-1)\times(n-1)$ sub-matrix of B, obtained by suppressing the first row and the first (non-null) column. Since the first co-ordinates of $\beta_2, \ldots, \beta_m$ are all null, the effect of any elementary row-operation on these row-vectors (of B) is determined by its effect on 2nd to nth co-ordinates, i.e., by the elementary row-operation *induced in* B_1.

$$B=\left[\begin{array}{c|c} 1 & * \cdots * \\ \hline 0 & \\ \vdots & B_1 \\ 0 & \end{array}\right]$$

$$C_1=\left[\begin{array}{c|c} 1 & * \cdots * \\ \hline 0 & \\ \vdots & C_2 \\ 0 & \end{array}\right]$$

If all the elements of B_1 be null, then B has the echelon form. When all the elements of B_1 are not null, the first non-null column-vector of B_1 is swept out (in the same way as in the case of A), and a matrix

C_1, which is row-equivalent to B_1, is obtained. Then the matrix C, obtained by replacing the sub matrix B_1 by C_1 in B, is row-equivalent to B, and hence also to A :

$$C=\left[\begin{array}{c|c} 1 & * \cdots * \\ \hline 0 & \\ \vdots & C_1 \\ 0 & \end{array}\right]=\left[\begin{array}{cc|c} 1 & * & * \cdots * \\ 0 & 1 & * \cdots * \\ \hline 0 & 0 & \\ \vdots & \vdots & C_2 \\ 0 & 0 & \end{array}\right],$$

The $(m-2)\times(n-2)$ matrix C_2, obtained by suppressing the first two rows and the first two (non-null) columns of C is the considered, and unless all the elements of C_2 happen to be null (and then C is in the echelon form) the above method is applied to C_2. Proceeding thus, one obtains in a finite number of steps, the desired echelon form E or a matrix which differs from E by a permutation of the columns (since the null columns in A, B_1, C_2, were disregarded in the enumeration of columns); the reduced matrix is row-equivalent to the given matrix A.

Every class of row-equivalent matrices is, by the above proposition, represented by a matrix in the echelon form. Hence echelon forms may be taken as normal forms of the matrices under row-equivalence.

Another type of normal form, called the *fully swept out form*, is obtained as follows.

Let the non-null row-vectors of the echelon form (1) be $\varepsilon_k=(e_{k1},\ldots,e_{kn})$, for $k=1,\ldots, r$, then

$$e_{ki}=0, \quad \text{for } i=1,\ldots, k-1, \quad e_{kk}=1,$$

and $\quad e_{ki}$ are arbitray elements, $\quad$ for $i=k+1,\ldots, n$.

Then the matrix S, obtained by replacing ε_k by the vector δ_k for $k=1,\ldots, r$, where $\delta_r=\varepsilon_r$, and

$$\delta_k=\varepsilon_k-e_{kk+1}\delta_{k+1}-\ldots-e_{kr}\delta_r, \quad \text{for } k=r-1,\ r-2,\ldots,\ 1,$$

is said to be in the fully swept out form.

Let $\delta_k=(d_{k1},\ldots, d_{kn})$, for $k=1,\ldots, r$, be the row-vectors of the matrix S, then $d_{kk}=1$, and

$d_{ki}=0$, for $i=1,\ldots, k-1, k+1,\ldots, r$,

d_{ki} are arbitrary elements for $i=r+1,\ldots, n$;

and the row-vectors δ_k are null vectors for $k=r+1,\ldots, m$,

Thus one obtains the following :

III.2.2. *Any matrix A, over a skew field K, is row-equivalent to a fully swept out matrix of the form :*

$$S=\left[\begin{array}{cccccccc} 1 & 0 & \ldots & 0 & * & \ldots & * \\ 0 & 1 & \ldots & 0 & * & \ldots & * \\ \ldots & & \ldots & & \ldots & & \ldots \\ 0 & 0 & \ldots & 1 & * & \ldots & * \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ \ldots & & \ldots & & \ldots & & \ldots \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \end{array}\right], \qquad \ldots(2)$$

or a matrix which differs from the above by a permutation of the columns only ; the asterisks denote arbitrary scalars.

Column-equivalence. As in the case of row-equivalence, one can define the column-equivalence of matrices, in terms of the three types of elementary column-operations on a matrix.

Analogues of the propositions III.2.1 and III.2.2 hold also for column-equivalence.

Equivalent matrices. A matrix B is said to be *equivalent* to a matrix A, over the skew-field K, if it can be obtained from A by some elementary row and column-operations. Thus row-equivalence and column-equivalence of matrices are particular cases of this equivalence.

For a given matrix, one can obtain a row-equivalent matrix in the fully swept out normal form (2) or a matrix differing from it by a permutation of columns. Now, by applying elementary column-operations (of the type I), the exact fully swept out form can be obtained ; and then by sweeping out the rows successively, one obtains a *fully reduced normal form* (3), as given below. In other words :

III.2.3. *Any matrix A, over a skew field K, is equivalent to a matrix in the fully reduced normal form :*

$$F=\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & & \cdots & & & \cdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & & \cdots & & & \cdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}. \qquad \ldots (3)$$

Exercises

1. Obtain the echelon normal form and the fully swept out normal form (or forms differing from these by permutation of columns) of the following matrices, under row-equivalence :

(i) $\begin{bmatrix} 2 & 3 & 4 & 1 \\ 4 & 6 & 5 & 11 \\ 1 & 1 & 1/2 & 5 \\ 6 & 7 & 6 & 21 \end{bmatrix}$, (ii) $\begin{bmatrix} 1 & -2 & 3 & 7 \\ 2 & 1 & -1 & 5 \\ 2 & -1 & 2 & 12 \\ 1 & 8 & -11 & -11 \end{bmatrix}$,

(iii) $\begin{bmatrix} 2 & 3 & 4 & 1 \\ 5 & 8 & 13 & 4 \\ 1 & 1 & -1 & 2 \\ 3 & 5 & 9 & 1 \end{bmatrix}$, (iv) $\begin{bmatrix} 1 & 1 & 1 & 2 \\ 2 & 1 & 3 & 1 \\ 1 & 3 & 2 & 5 \\ 3 & -2 & 1 & k \end{bmatrix}$.

2. Find which of the above matrices are (a) row equivalent, (b) column-equivalent, (c) equivalent.

3. Row and column spaces of a matrix.

Let $A=(a_{ij})$ be an $m\times n$ matrix, over a skew field K. The linear space $R(A)$, generated by the row-vectors α_1, α_2, ..., α_m, is called the *row-space* of A. Likewise the linear space $C(A)$, generated column-vectors $\overset{(1)}{\alpha}$, $\overset{(2)}{\alpha}$, .., $\overset{(n)}{\alpha}$, is the *column-space* of A. $R(A)$ is a sub-space of $V_n(K)$ and $C(A)$ is a sub-space of $V_m(K)$.

III.3.1. *Two row-equivalent matrices have the same row-space.*

Proof: Let A and B be two row-equivalent matrices. The row-space $R(A)$ is generated by the row-vectors $\alpha_1, \ldots, \alpha_m$. It follows, by the replacement theorem II.1.8, that an elementary row-operation (of any type) transforms the set $\alpha_1, \ldots, \alpha_m$ to another set of generators of the same linear space $R(A)$. Hence $R(B)$ is the same as $R(A)$.

III.3.2. *Column-spaces of two row-equivalent matrices have the same rank.*

Proof: Let A and B be two row-equivalent matrices. $C(A)$ is generated by the column-vectors $\alpha^{(1)}, \ldots, \alpha^{(n)}$. Any relation of linear dependence among the column vectors,

$$g_1\alpha^{(1)} + \ldots + g_n\alpha^{(n)} = \theta,$$

is equivalent to the following system of equations in terms of the co-ordinates of the vectors :

$$g_1a_{k1} + \ldots + g_na_{kn} = 0, \qquad \text{for } k = 1, 2, \ldots, m.$$

Thus $g_1, \ldots, g_n$, the coefficients of the linear dependence, constitute a solution of the following system of equations :

$$x_1a_{k1} + \ldots + x_na_{kn} = 0, \quad \text{for } k = 1, \ldots, m. \qquad \ldots (1)$$

Now, (I) by an interchange of any two row vectors, the system of equations (1) remains unchanged.

(II) By replacing a row-vector α_k by the vector $c\alpha_k$, where $c \neq 0$, the kth equation of the system (1) is replaced by the equation

$$x_1 . ca_{k1} + \ldots. + x_n\, ca_{kn} = 0,$$

the other eqations remaining unaltered ; thus the system (1) is transformed to a system, equivalent to it.

(III) By replacing α_r by $\alpha_r + c\alpha_k$ the rth equation of the system (1) is replaced by the equation

$$x_1(a_{r1} + ca_{k1}) + \ldots + x_n(a_{rn} + ca_{kn}) = 0$$

which together with the kth equation, is eqivalent to the pair of the rth and kth equations. Since all equations, other than the rth one remains unaltered by this operation, it follows that this operation transforms the system (1) to an equivalent system of equations.

Thus a linear dependence of some column-vectors of A implies a linear dependence of the corresponding column-vectors of a matrix obtained by applying any elementary row-operation on A. Again, since inverse of an elementary row-operation is also an elementary row-operation (of the same type), a linear dependence of certain column-vectors cannot hold after an elementary row-operation unless the corresponding column-vectors were linearly dependent before the operation.

Let rank of $C(A)$ be s, then s suitable column-vectors of A form a basis of $C(A)$; these are linearly independent, and each of the remaining $n-s$ column-vectors is linearly dependent on these. After applying any elementary row-operation on A, the column vectors, corresponding to the above basis vectors of $C(A)$ continue to be linearly independent, and each of the remaining $n-s$ column-vectors are linearly dependent on these. Hence the rank of $C(A)$ remains unaltered by any elementary row-operation on A. Thus rank of $C(A)=$rank of $C(B)$.

It should however be noted that the column-space of a matrix does *not* remain unaltered by an elementary row-operation (though the rank of column-space remains the same).

Rank of a matrix. One finally obtains the following main theorem :

III.3.3. *For any matrix A, rank$[R(A)]=$rank $[C(A)]$.*

Proof : The matrix A is row-equivalent to a fully swept out matrix S of the form (2), page 65, (or a matrix which differs from S by a permutation of the columns). Then, it follows from III.3.1 and III.3.2 that

$$\text{rank }[R(A)]=\text{rank }[R(S)],\quad \text{rank }[C(A)]=\text{rank }[C(S)]$$

The non-null row-vectors $\delta_1,\ldots,\delta_r$ of the matrix S generate $R(S)$; these vectors are also linearly independent. Hence rank $[R(S)]=r$. Again, the first r (non-null) column-vectors of S are linearly independent; and since every column-vector is linearly dependent on these, rank $[C(S)]=r$.

Hence rank $[R(A)]=r=$rank $[C(A)]$.

Rank of $R(A)$, which is the same as the rank of $C(A)$, is called the *rank of the matrix* A.

EXAMPLE 1: The method of sweep out has been applied in example 2, p. 89. Taking k, μ, ϕ, ρ as the row-vectors of a matrix A, it follows from that example that the row-equivalent matrix P, to which A is finally reduced, differs from the fully swept out normal form S by a permutation with carries the second column to the last position.

$$P=\begin{bmatrix} 1 & -2 & 0 & 0 & -800/199 \\ 0 & 0 & 1 & 0 & 80/199 \\ 0 & 0 & 0 & 1 & -282/199 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad S=\begin{bmatrix} 1 & 0 & 0 & -300/199 & -2 \\ 0 & 1 & 0 & 80/199 & 0 \\ 0 & 0 & 1 & -282/199 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix};$$

rank of the matrix A is 3.

Since rank of a matrix remains unaltered by any elementary row or column operation (by III.3.1 and III.3.2), it follows that:

✓III.3 4. *Two equivalent matrices have the same rank.*

Let E, S and F, as given in III.2.1, III.2.2, and III.2.3, be the matrices in the echelon form, fully swept out form and fully reduced normal form, such that E and S are row-equivalent to A, and F is equivalent to A, where A is a given $m \times n$ matrix. Then rank (E) = rank (S) = rank (F) = rank $(A) = r$ (say).

Also ranks of E, S and F are given be the number of their non-null row vectors (as these row vectors are linearly independent, and genarate their respective row-spaces). Hence one obtains the following:

III.3 5. *Any $m \times n$ matrix of rank r is*

(*i*) *row-equivalent to an echelon matrix E with r non-null rows,*

(*ii*) *row equivalent to a full swept out matrix S with r non-null rows,*

(*iii*) *equivalent to a matrix F, in the fully reduced normal form with r identity elements in its principal diagonal.*

As a matrix F, in the fully reduced normal form, is determined uniquely by the number of identity elements (in its principal diagonal), it follows from III.3.4 and III.3.5 (*iii*) that

III.3.6, *Two $m \times n$ matrices are equivalent if, and only if, they have the same rank.*

Relations between the ranks of matrices and the ranks of their sum and products are obtained in the following proposition.

III.3.7. *If A and B be two $m\times n$ matrices, and G any $n\times k$ matrix, then*

(i) rank $(A+B)\leqslant$ rank $(A)+$rank (B),

(ii) rank $(A.G)\leqslant$ minimum [rank (A), rank (G)].

Proof: (*i*) Let $\alpha_1,\dots,\alpha_m$ and $\beta_1,\dots,\beta_m$ be the row-vectors of the matrices A and B, then the row vectors of their sum $A+B$ are $\alpha_1+\beta_1,\dots,\alpha_m+\beta_m$. These three sets of row-vectors generate $R(A)$, $R(B)$ and $R(A+B)$, the row-spaces of the matrices A, B and $A+B$ respectively The vectors $\alpha_1,\dots,\alpha_m,\beta_1,\dots,\beta_m$ generate the linear sum $R(A)+R(B)$ of the spaces $R(A)$ and $R(B)$. Since the vectors $\alpha_1+\beta_1,\dots,\alpha_m+\beta_m$ lie in the space $R(A)+R(B)$, it follows that $R(A+B)$ is a sub-space of $R(A)+R(B)$. Hence rank $[R(A+B)]\leqslant$ rank $[R(A)+R(B)]$. But rank $[R(A)+R(B)]\leqslant$ rank $(R(A))+$rank $(R(B))$ (by II.2 5). Also, since rank of a matrix is the same as the rank of its row-space, it follows that rank $(A+B)\leqslant$ rank $(A)+$rank (B).

(ii) Let $A=(a_{ij})$ and $G=(g_{ij})$; and let $\alpha^{(1)},\dots,\alpha^{(n)}$ be the column-vectors of A, and $\gamma_1,\dots,\gamma_n$ be the row-vectors of G. Let $P=(p_{ji})$ be the product matrix, $P=AG$, and let $\pi_1,\dots,\pi_m$ be the row vectors and $\pi^{(1)},\dots,\pi^{(k)}$ be the column-vectors of P. Then for $i=1,\dots,m$,

$$\begin{aligned}\pi_i=(p_{i1},\dots,p_{ik})&=(a_{i1}g_{11}+\dots+a_{in}g_{n1},\dots,a_{i1}g_{1k}+\dots\dots a_{in}g_{nk})\\&=a_{i1}(g_{11},\dots,g_{1k})+\dots\dots+a_{in}(g_{n1},\dots,g_{nk})\\&=a_{i1}\gamma_1+\dots+a_{in}\gamma_n\in R(G),\end{aligned}$$

so that $R(P)$ is a sub-space of $R(G)$, where $R(G)$, $R(P)$ are the row-spaces of the matrices G and P. Hence rank $(R(P))\leqslant$ rank $(R(G))$, and therefore rank $(AG)=$rank $(P)\leqslant$ rank (G).

Similarly, for $i=1,\dots,k$,

$$\pi^{(i)}=(p_{1i},\dots,p_{mi})=\alpha^{(1)}g_{1i}+\dots+\alpha^{(n)}g_{ni}\in C(A),$$

so that $C(P)$ is a sub-space of $C(A)$, where $C(A)$, $C(P)$ are the column-spaces of A and P. Then rank $(C(P))\leqslant$rank $(C(A))$, from which, it follows that rank $(AG)=$rank $(P)=$rank $(C(P))\leqslant$ rank $(C(A))=$rank (A).

Combining the above two results, one gets,

$$\text{rank } (A.G) \leqslant \text{minimum } [\text{ rank } (A), \text{ rank } (G)].$$

EXAMPLE 2 : An $m \times n$ matrix A has rank 1 if, and only if, $A=BC$. where B and C are $m \times 1$ and $1 \times n$ non-null matrices respectively.

Let $A=BC$, where B and C are $m \times 1$ and $1 \times n$ non-null matrices respectively. Then rank $(B)=$rank $(C)=1$, and A is an $m \times n$ matrix, such that rank $(A)=$rank$(BC) \leqslant 1$, by III.3.7 (ii). Hence rank$(A)=1$, as A is necessarily a non-null matrix.

Conversely, let $A=(a_{ij})$ be an $m \times n$ matrix of rank 1. Since the row-space of A is of rank 1, there exists at least one non-null row-vector, say $\alpha_r=(a_{r1}, \ldots, a_{rn})$, and all the row-vectors are linearly dependent on it. Thus

$\alpha_i=(u_i a_{r1}, \ldots, u_i a_{rn})$, for $i=1, \ldots, m$, with $u_r=1$: consequently,

$$A=(a_{ij})=\begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix}(a_{r1}, \ldots, a_{rn})=BC,$$

where B is an $m \times 1$ and C a $1 \times n$ matrices. Also, rank $(B)=1$, as $u_r \neq 0$; and rank $(C)=1$, as α_r is a non-null vector,

Nullity of a matrix. *Nullities of an $m \times n$ matrix A are* defined by : row nullity$(A)=m-$rank(A), and column nullity (A) $=n-$rank(A). For an $n \times n$ matrix A, the two nullities are equal, and nullity $(A)=n-$rank(A). One obtains properties of nullities of matrices from different properties of their ranks ; viz. from III.3.7, one obtains :

III.3.8. *If A and B be two $n \times n$ matrices, then*

(i) *nullity $(A+B) \geqslant$ nullity (A) + nullity $(B)-n$, and*

(ii) *nullity $(A.B) \geqslant$ maximum [nullity (A), nullity (B)].*

The following important result on nullities of matrices, known as *Sylvester's law of nullity*, will be proved later :

III.3.9. *If A and B be two $n \times n$ matrices, then*

(i) *nullity $(A)+$nullity $(B) \leqslant$ nullity $(A.B)$, and*

(ii) *rank $(A)+$rank $(B)-n \leqslant$rank $(A.B)$.*

III.3.9 (ii) is the dual of the property III.3.8 i) for nullities.

Exercises

1. Determine the ranks of the matrices given in exercise 1, page 66, and in exercise 5 (below),

(ii) by finding the rank of their row-spaces ;

(ii) by finding the rank of their column-spaces ;

(iii) by reducing them to their fully reduced normal forms.

2. Prove that the rank of the following matrix is 0 or 3 :

$$A=\begin{bmatrix} 0 & 0 & 0 & b & c & d \\ 0 & -d & c & -a & 0 & 0 \\ d & 0 & -b & 0 & -a & 0 \\ -c & b & 0 & 0 & 0 & -a \end{bmatrix}.$$

If $\alpha^{(1)},\ldots,\alpha^{(6)}$ be the column-vectors of A, and $y_1\alpha^{(1)}+\ldots+y_6\alpha^{(6)}=(0,0,0,0)$, then $y_1y_4+y_2y_5+y_3y_6=0$.

3. If A be an $m\times n$ matrix of rank r, then $A=BC$, where B and C are $m\times k$ and $k\times n$ matrices of rank r (but B and C are not unique).

4. The rank of a matrix is not changed if the field of scalars be enlarged.

5. Find the nullities of the following matrices :

$$\begin{bmatrix} 2 & -3 & 5 & 7 \\ -1 & 2 & -3 & -2 \\ 1 & -1 & 2 & 3 \end{bmatrix}, \quad \begin{bmatrix} 1 & -2 & 3 & 4 & 5 \\ -3 & 2 & -1 & -3 & 4 \\ 2 & 0 & -2 & 7 & -9 \end{bmatrix}$$

4. Square matrices.

The product of two matrices exists only for a conformable pair ; hence a set of matricess can be closed with respect to multiplication if, and only if, the matrices be all square matrices of the same size. An $n\times n$ matrix is said to be of *order n*.

All matrices treated in this article are square matrices of order n over a *field K*, unless otherwise stated. The set of all such matrices will be denoted by $\Lambda_n(K)$, or simply Λ_n.

Total matric algebra Λ_n is a linear space of rank n^2 over K (by III.1.2). Further, Λ_n is closed with respect to multiplication, and multiplication is associative and distributive (by III.1.3). Also the matrix I, whose elements in the principal diagonal are all indentity and all other elements are null, satisfies $AI=IA=A$ for every $A \epsilon \Lambda_n$.

$$I=\begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & 1 \end{pmatrix}.$$

Thus Λ_n is a ring with the identity element I.

Also, since multiplication is commutative in K, it follows that for any $k \epsilon K$, and $A, B \epsilon \Lambda_n$ $(kA)B=A(kB)=k(AB)$. By virtue of these properties :

III.4.1. *The set Λ_n of all square matrices of order n, over a field K, forms an algebra, called the total matric algebra.*

This belongs to a class of more general systems, called *algebras* (or *hypercomplex number systems*).

Some special square matrices, of considerable importance, will now be introduced.

Diagonal matrices. A matrix $D=(d_{ij})$ is called a *diagonal matrix* if $d_{ij}=0$ when $i \neq j$, and $d_{ii}=d_i$ for $i=1, 2, \ldots, n$; the elements $d_1, \ldots, d_n$, some of which may also by null, are called the *diagonal elements* of D ; and D is denoted by (d_i).

From the law of multiplication of matrices, it is evident that :

The effect of multiplying a matrix A from the *left* by a diagonal matrix $D=(d_i)$ is to multiply the ith *row-vector* of A by d_i, for $i=1, 2, \ldots, n$; and the effect of multiplying a matrix A from the *right* by a diagonal matrix $D=(d_i)$ is to multiply the ith *column-vector* of A by d_i, for $i=1, 2, \ldots, n$.

Scalar matrices. When the diagonal elements of a diagonal matrix D are all equal, say $d_i=d$ for $i=1, 2, \ldots, n$, D is called the *scalar-matrix representing* d, and is denoted by (d). The scalar matrix (1) is the identity matrix I, and the scalar matrix (0) is the null matrix O.

The product of a scalar matrix (d) and any arbitrary matrix A is the matrix whose elements are all d times the corresponding elements of A ; hence it is the same as the matrix $d.A$, the product of A by the scalar d. In particular, $(d)=d.I$. Also the product of two scalar matrices $(c).(d)=(cd)$ is a scalar matrix; and the inverse of (d) is the scalar matrix (d^{-1}) which is uniquely determined so long as $d\neq 0$. The sum of the two scalar matrices corresponding to c and d is the scalar matrix corresponding to the sum $c+d$, *i.e.*, $(c)+(d)=(c+d)$. Thus one obtains the following :

III.4.2. *The scalar matrices form a subfield of the ring $\Lambda_n(K)$ (consisting of all square matrices of order n), isomorphic to the field of scalars K.*

Elementary matrices. A matrix obtained by applying any *one* of the elementary row-operations (page 62) on the indentity matrix I is called an *elementary matrix*. Thus there are three different types of elementary matrices :

(I) the elementary matrix obtained by interchanging the ith and the kth rows of I, being denoted by I_{ik} ;

(II) the elementary matrix obtained by multiplying the kth row by a non-unll scalar c, being denoted by $I_k(c)$;

(III) the elementary matrix obtained by replacing the ith row by the sum of the ith row and c-times the kth row, being denoted by $I_{ik}(c)$.

The effect of multiplying any arbitrary matrix A by any of these three types of elementary matrices from the *left hand side* is to apply the corresponding *row-operation* on A. These are best explained by the following illustrations :

$$I_{23}.A=\begin{bmatrix}1&0&0\\0&0&1\\0&1&0\end{bmatrix}.\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}=\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{31}&a_{32}&a_{33}\\a_{21}&a_{22}&a_{23}\end{bmatrix};$$

the second and third rows have been interchanged.

$$I_2(c).A=\begin{bmatrix}1&0&0\\0&c&0\\0&0&1\end{bmatrix}.\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}=\begin{bmatrix}a_{11}&a_{12}&a_{13}\\ca_{21}&ca_{22}&ca_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}$$

the second row has been multiplied by the scalar c.

$$I_{23}(c).A=\begin{bmatrix}1&0&0\\0&1&c\\0&0&1\end{bmatrix}.\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}$$

$$=\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}+ca_{31}&a_{22}+ca_{32}&a_{23}+ca_{33}\\a_{31}&a_{32}&a_{33}\end{bmatrix};$$

the second row has been replaced by the sum of the second row and c-times the third row.

Any elementary matrix of any of the three types is also obtained by applying a corresponding column-operation on the identity matrix, and the effect of multiplying any matrix A by any elementary matrix from the *right hand side* is to apply the corresponding *column-operation* on A.

It may however be noted that the elementary matrix $I_{ik}(c)$ is obtained from I either by adding c-times the kth row to to the ith row, or by adding c-times the ith column to the kth column. If a matrix A be multipled from the left hand side by $I_{ik}(c)$, c-times the kth row of A is added to its ith row, while if A is multiplied form the right hand side by $I_{ik}(c)$, c-times the ith column of A is added to its kth column.

It has been proved in III.2.3 that any matrix A is equivalent to a matrix F of the form (3), page 66. F has been obtained by applying some elementary row and column-operations on A. Hence it follows that :

III.4.3. *For any given $m\times n$ matrix A there exists an $m\times n$ matrix F in the fully reduced normal form* (3), *page* 6 , *such that*

$$F=E_1\dots E_rAE_{r+1}\dots E_m,$$

where $E_1,\dots,E_m$ are suitable elementary matrices.

Non-singular matrices. A matrix R is said to be a *right inverse* of an $m\times n$ matrix A if $AR=I$; likewise L is a *left inverse* of A if $LA=I$. Since I is a square matrix, R and L are necessarily $n\times m$ matrices. A right (left) inverse may not be unique, and it may not even exist (see exercise 1, page 83).

A matrix B is said to be a *both-sided inverse* or simply an *inverse* of a given matrix A, if $AB=I=BA$. If A be an $m\times n$

matrix, B must be an $n\times m$ matrix. Then AB and BA will be respectively $m\times m$ and $n\times n$ matrices ; and since these two are equal, it follows that $m=n$. Thus a matrix having a (both-sided) inverse is necessarily a square matrix.

Again, if a square matrix A has a right inverse R and a left inverse L, then $R=L$; in fact,

$$R=IR=(LA)R=L(AR)=LI=L.$$

It will be shown later that if a square matrix has a right (left) inverse, then the same is also a left (right) inverse, and it is the *unique* both-sided inverse.

A square matrix, for which an inverse matrix exists, is called *non-singular* ; otherwise it is said to be *singular*.

In particular, every elementary matrix has an inverse. In fact, it follows from the rules of multiplication by elementary matrices that.

$$I_{ij}.I_{ij}=I, \qquad i.e.. \quad I_{ij}^{-1}=I_{ij} ;$$

$$I_k(c).I_k(c^{-1})=I=I_k(c^{-1}).I_k(c), \qquad i.e., \quad [I_k(c)]^{-1}=I_k(c^{-1}) ;$$

and $$I_{ik}(c).I_{ik}(-c)=I=I_{ik}(-c).I_{ik}(c), \qquad i.e, \quad [I_{ik}(c)]^{-1}=I_{ik}(-c).$$

Thus, one obtains :

III.4.4. *Every elementary matrix is non-singular, the inverse of an elementrry matrix being an elementary matrix of the same type.*

Again, for a diagonal matrix $D=(d_i)$ and a matrix A, $DA=I=AD$ implies that A must be the diagonal matrix (d_i^{-1}) ; thus

III.4.5. *A diagonal matrix is non-singular if, and only if, the diagonal elements are all non-null. The inverse of the diagonal matrix* (d_i) *is the diagonal matrix* (d_i^{-1}).

It follows from III.4.3 that for a given square matrix A of order n, there exists a diagonal matrix F of order n whose diagonal elements are 1 or 0, and such that

$$F=E_1\ldots E_rAE_{r+1}\ldots E_m, \qquad \ldots (1)$$

where $E_1,\ldots,E_m$ are suitable elementary matrices. Since every elementary matrix has an inverse, it follows that

$$A=E_r^{-1}\ldots E_1^{-1}FE_m^{-1}\ldots E_{r+1}^{-1}. \qquad \ldots (2)$$

Now, if A^{-1} exists, then F^{-1} also exists. and is given by

$$F^{-1}=E_m^{-1}\ldots E_{r+1}^{-1}A^{-1}E_r^{-1}\ldots E_1^{-1}.$$

Conversely, when F^{-1} exists, then A^{-1} also exists, and

$$A^{-1}=E_{r+1}\ldots E_m F^{-1} E_1 \ldots E_r.$$

Again, since F is a diagonal matrix, whose diagonal elements are *1* or *0*, F^{-1} exists if, and only if, all its diagonal elements are non-null (by III.4.5), i.e., when all the diagonal elements are *1*, so that $F=I$. Thus

III.4 6. *A matrix A is non-singular if, and only if, it is equivalent to the identity matrix I.*

Now, since the inverse of an elementary matrix is an elementary matrix, it follows from III.4.6 and (2) that

III.4.7. *A matrix is non-singular if, and only if, it is a product of some elementary matrices.*

The same non-singular matrix may however be expressed as products of different sets of elementary matrices (see example 1, page 78).

As an immediate consequence III.4.7 and III.4.4 it follows that product of two non-singular matrices is a non-singular matrix, and the inverse of a non singular matrix is also a non-singular matrix. Hence

III.4 8. *All non-singular matrices of order n, over a field K, form a group, called the full (or general) linar group $GL(n,K)$.*

Since pre-multiplication, *i e.*, multiplication from left hand side, of a matrix M (which may, in general, be rectangular) by an elementary matrix means an elementary, row-operation on M, and post-multipication *i.e.*, multiplication from right hand side, of M by an elementary matrix means an elementary column-operation on M, and conversely, it follows that *a matrix N is equivalent to M if, and only if $N=PMQ$, where P,Q are non-singular matrics.* Also since the rank of M remains unaltered by any row or column-operation on M (by III.3.4), it follows from III.4.7 that

III.4.9. *The rank of a matrix remains unaltered on multiplication by any non-singular matrix (from either side).*

Again, since the rank of the identity matrix I, of order n, is n, it follows by III.4.6 that

III.4.10. (i) *A matrix A is non-singular if, and only if, its rank is equal to its order.*

(ii) *A matrix A is non-singular if, and only if, the row-vector (also the column vectors) of A are linearly independent.*

EXAMPLE 1 : To find whether the following matrices are non-singular,

$$\text{(i)}\quad A=\begin{bmatrix}1 & 2 & -1\\ 2 & -3 & 1\\ 3 & -1 & 1\end{bmatrix}\qquad \text{(ii)}\quad B=\begin{bmatrix}2 & -3 & -2\\ 1 & 2 & 1\\ 3 & -1 & -1\end{bmatrix}$$

and to express the non-singular matrices (if any) as products of elementary matrices.

(i) By applying row-operations only, one obtains

$$I_{31}(-3)I_{21}(-2)A=\begin{bmatrix}1 & 2 & -1\\ 0 & -7 & 3\\ 0 & -7 & 4\end{bmatrix},$$

$$\rightarrow\begin{bmatrix}1 & 2 & -1\\ 0 & 1 & -3/7\\ 0 & -7 & 4\end{bmatrix}$$ on pre-multiplication by $I_3(-1/7)$,

$$\rightarrow\begin{bmatrix}1 & 0 & -1/7\\ 0 & 1 & -3/7\\ 0 & 0 & 1\end{bmatrix}$$ on pre-multiplication by $I_{32}(7).I_{12}(-2)$.

$$\rightarrow\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$$ on pre-multiplication by $I_{23}(3/7)\, I_{13}(1/7)$.

$$I_{13}(3/7)I_{13}(1/7)I_{32}(7)I_{12}(-2)I_2(-1/7)I_{31}(-3)I_{21}(-2).A=I.$$

$$A=[I_{23}(3/7)I_{13}(1/7)I_{32}(7)I_{12}(-2)I_2(-1/7)I_{31}(-3)I_{21}(-2)]^{-1}$$

$$=I_{21}(2)I_{31}(3)I_2(-7)I_{12}(2)I_{32}(-7)I_{13}(-1/7)I_{23}(-3/7).$$

The matrix A, being a product of elementary matrices, is non-singular.

Alternatively : By applying row-operations

$$I_{32}(7)I_2(-1/7)I_{31}(-3)I_{21}(-2).A=\begin{bmatrix}1 & 2 & -1\\ 0 & 1 & -3/7\\ 0 & 0 & 1\end{bmatrix}$$

and then by applying column-oprations,

$$\begin{bmatrix}1 & 2 & -1\\ 0 & 1 & -3/7\\ 0 & 0 & 1\end{bmatrix}.I_{12}(-2)I_{13}(1)I_{23}(3/7)=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}=I.\,,$$

Thus $I_{32}(7)I_2(-1/7)I_{31}(-3)I_{21}(-2).A.\ I_{12}(-2)I_{13}(1)I_{23}(3/7)=I$,

$A=I_{21}(2)I_{31}(3)I_2(-7)I_{32}(-7)I_{32}(-3/7)I_{13}(-1)I_{12}(2)$.

Consequently, A is non-singular.

Also, A has two different representations as products of elementary matrices.

(ii) Applying row-operations on B,

$I_{32}(-7/2)I_3(2/7)I_{31}(-3)I_{21}(-1)I_1(1/2).B=C$;

then applying column-operations,

$$C.I_{12}(3/2)I_{13}(1)I_{23}(-4/7)=\begin{bmatrix}1&0&0\\0&1&0\\0&0&0\end{bmatrix}=F.$$

Since rank $(F)=2\neq 3$, B is singular.

Symmetric and skew-symmetric matrices. A matrix $A=(a_{ik})$ is said to be *symmetric* if $A^t=A$, i.e., $a_{ki}=a_{ik}$ for i, $k=1, 2, \ldots, n$. Diagonal matrices and elementary matrices of the types I_{ik}, $I_k(c)$ are symmetric matrices.

A matrix $A=(a_{ik})$ is said to be *skew-symmetric, anti-symmetric* or *alternating* if $A^t=-A$. Thus $a_{ki}=-a_{ik}$ for i, $k=1,2,\ldots,n$, in the case of a skew-symmetric matrix. In particular, $a_{kk}=-a_{kk}$, i.e., $2a_{kk}=0$; and when the characteristic of the ground field K is different from 2, $a_{kk}=0$, for $k=1, 2,\ldots,n$, Thus

III.4.11. *The diagonal elements of a skew-symmetric matrix, over a field of characteristic* $\neq 2$, *are all null.*

For an arbitrary square matrix, one obtains :

III.4.12. *Any matrix can be expressed as a sum of a symmetric and a skew-symmetric matrix.*

Proof : Let $A=(a_{ik})$ be a given matrix, and let

$S=\frac{1}{2}(A+A^t)$ and $Q=\frac{1}{2}(A-A^t)$, then $A=S+Q$,

where S is symmetric and Q skew-symmetric. In fact,

$$S^t=\tfrac{1}{2}(A+A^t)^t=\tfrac{1}{2}(A^t+A^{tt})=\tfrac{1}{2}(A^t+A)=S,$$

and $$Q^t=\tfrac{1}{2}(A-A^t)^t=\tfrac{1}{2}(A^t-A^{tt})=\tfrac{1}{2}(A^t-A)=-Q.$$

For any $m\times n$ matrix A, the products AA^t and A^tA are square matrices of orders m and n respectively, and both are symmetric matrices, since

$$(AA^t)^t=A^{tt}A^t=AA^t \text{ and } (A^tA)^t=A^tA^{tt}=A^tA.$$

EXAMPLE 2 : For a 2×3 matrix A, one obtains as follows :

$$\begin{pmatrix}2&-1&0\\3&2&-4\end{pmatrix}\begin{bmatrix}2&3\\-1&2\\0&-4\end{bmatrix}=\begin{pmatrix}5&4\\4&29\end{pmatrix}.$$

$$\begin{bmatrix}2&3\\-1&2\\0&-4\end{bmatrix}\begin{pmatrix}2&-1&0\\3&2&-4\end{pmatrix}=\begin{bmatrix}13&4&-12\\4&5&-8\\-12&-8&16\end{bmatrix}.$$

The matrices AA^t and A^tA are not the same, in general, even when A is itself a square matrix, e.g.

$$\begin{pmatrix}2&5\\3&-1\end{pmatrix}\begin{pmatrix}2&3\\5&-1\end{pmatrix}=\begin{pmatrix}29&1\\&10\end{pmatrix},\ \begin{pmatrix}2&3\\5&-1\end{pmatrix}\begin{pmatrix}2&5\\3&-1\end{pmatrix}=\begin{pmatrix}13&7\\2&25\end{pmatrix}$$

Orthogonal matrices. A matrix $P=(p_{ik})$ is said to be an *orthogonal matrix*, if $PP^t=I$

Since P is a square matrix and P^t is a right inverse of P, it is the unique both-sided inverse of P, so that $P^t=P^{-1}$. Now, from $P^t(P^t)^t=P^tP=P^{-1}P=I$, it follows that

III.4.13. *The transposed of an orthogonal matrix is its inverse, and it is also orthogonal.*

As the inverse of an orthogonal matrix exists (being its own transposed), every othogonal matrix is non-singular. Moreover, if A, B be two orthogonal matrices then the product AB is also orthogonal, since

$$(AB)(AB)^t=(AB)(B^tA^t)=A(BB^t)A^t=AIA^t=AA^t=I.$$

The inverse matrix is also an orthogonal matrix.

Thus one obtains :

III.4.14. *The orthogonal matrices of order n, over a field K, form a sub-group of the full linear group $GL(n, K)$.*

The group of the orthogonal matrices is called *the orthogonal group* $O(n, K)$.

III 4.15. *A real matrix of order n, is orthogonal if, and only if, its row-vectors form an orthonormal system in $V_n(E)$.*

Proof : Let P be a real orthogonal matrix. Let π_i be the ith row-vector of P ; then it is the ith column-vector of P^t, for $i=1,2,\dots, n$. Again the elements of I are *Kronecker deltas*

δ_{ik} (where $\delta_{ii}=1$ and $\delta_{ik}=0$ when $i\neq k$), for $i, k=1, 2,\ldots, n$. Then, it follows from the relation $PP^t=I$ that

$$\pi_i\pi_k=(i\text{th row-vector of } P).(k\text{th column-vector of } P^t)$$
$$=\delta_{ik}.$$

Conversely, $\pi_i\pi_k=\delta_{ik}$ implies $PP^t=I$, so that P is an orthogonal matrix.

Since the row-vectors of P^t are the column-vectors of P, it follows, from IV.4.15, that

III.4.16. *A real matrix, of order n, is orthogonal if, and only if, its column-vectors form an orthonormal system in $V_n(E)$.*

EXAMPLE 8: An orthogonal matrix of order 2, over the field of real numbers, is of either of the two forms;

$$R=\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}, \quad S=\begin{pmatrix}\cos\theta & \sin\theta\\ \sin\theta & -\cos\theta\end{pmatrix}.$$

Let $G=\begin{pmatrix}a & b\\ c & d\end{pmatrix}$ be an orthogonal matrix, then

$$a^2+b^2=1,\ c^2+d^2=1, \text{ and } ac+bd=0.$$

Putting $a=\cos\phi$, $c=\sin\psi$, one gets $b=\pm\sin\phi$ and $d=\pm\cos\psi$. Then from $0=ac+bd=\cos\phi\sin\psi\pm\sin\phi\cos\psi=\sin(\psi\pm\phi)$, one obtains $\psi=\pm\phi$, so that the following different cases arise:

$b=\sin\phi$, $d=\cos\psi$, so that $\psi=-\phi$, and $G=R$ for $\phi=-\theta$.
$b=\sin\phi$, $d=-\cos\psi$, so that $\psi=\phi$, and $G=S$ for $\phi=\theta$.
$b=-\sin\phi$, $d=\cos\psi$, so that $\psi=\phi$, and $G=R$ for $\phi=\theta$.
$b=-\sin\phi$, $d=-\cos\psi$, so that $\psi=-\phi$, and $G=S$ for $\phi=-\theta$.

Exercises

1. A scalar matrix is commutative with every matrix.

2. (i) Find the set of all matrices which are commutative with a given diagonal matrix.

(ii) If a matrix be commutative with all diagonal matrices of order n, then it is necessarily itself a diagonal matrix.

3. The non-singular diagonal matrices form a sub-group of the full linear group $GL(n,K)$.

4. [The matrices I_{km} together with I are called *permutation*

matrices. A matrix I_{km} multiplied by some matrices of the type $I_r(c)$ is called a *monomial matrix*].

(i) Show that a monomial matrix M can always be expressed as $M=DP$, where D is a non-singular diagonal matrix and P a permutation matrix.

(ii) Let P, D, M be respectively a permutation diagonal and a monomial matrix ; find the nature of the matrices PDP^{-1}, MDM^{-1}.

5. [*A matrix* A, of order n, is said to be *triangular* (or *upper triangular*, in particular), if $a_{ij}=0$, for $i>j$, *and* $i, j=1, 2, \ldots n$. If, moreover, $a_{ii}=0$, for $=1, \ldots\ldots, n$, then it is said to be *strictly triangular.*]

(i) The sum and product of two triangular matrices (of the same order) is triangular.

(ii) A triangular matrix is non-singular if, and only if, all its diagonal elements be non-null.

(iii) The set of all non-singular triangular matrices (of order n) froms a subgroup of $GL(n,K)$.

6. [A square matrix A is said to be *nilpotent of index* s, if $A^s=O$, and $A^{s-1}\neq O$.]

Show that any strictly triangular matrix is nilpotent.

7. If P be a singular matrix of order n, then $P^2=c.P$, where c is a scalar, for $n=1,2$ (the same may not be true for $n>2$).

8. The set of all real matrices $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ form a field isomorphic to the field of complex numbers.

9. The set of all matrices,

$$\begin{pmatrix} a & -b & -c & -d \\ b & a & -d & -c \\ c & d & a & b \\ -d & c & -b & a \end{pmatrix}$$

over the rational field, forms a skew field, isomorphic to the skew field formed by the the rational quaternions.

10. If A and B be two equivalent matrices, find whether the following pairs of matrices are equivalent :

(i) A^t and B^t, (ii) A^2 and B^2, (iii) AB and BA.

11. (i) Find a right inverse of the matrix $\begin{pmatrix} 2 & -3 & 1 \\ 5 & -1 & 2 \end{pmatrix}$. Show that it has infinitely many right inverses, but no left inverse.

(ii) Find a matrix which has infinitely many left inverses, but no right inverse.

12. [A matrix P is said to be *idempotent*, if $P^2 = P$. The null matrix and the identity matrix I are idempotent ; there are others ; e. g. the matrix $\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$ is idempotent].

Show that an idempotent matrix, other than I, is singular.

13. If AB be non-singular, then each of A and B is non-singular.

14. Find the order of the group $GL(2, GF(p))$.

15. Express the following as products of elementary matrices (where possible) ; then find their inverse matrices :

$$\begin{pmatrix} 7 & 3 \\ 5 & 2 \end{pmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 2 & -6 & 1 \\ 3 & 4 & 2 \end{bmatrix}, \begin{bmatrix} 6 & 4 & 3 \\ 1 & 2 & 3 \\ 1 & -2 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 3 & 7 \\ 2 & -1 & 5 \\ 1 & -4 & -2 \end{bmatrix}.$$

16. (i) Express each of the matrices of exercise 15 as a sum of a symmetric and a skew-symmetric matrix

(ii) The matrix $\begin{pmatrix} (1) & (1) \\ (0) & (1) \end{pmatrix}$, over the field $GF(2)$, cannot be expressed as a sum of a symmetric and a skew-symmetric matrix.

17. If a symmetric matrix has rank 1, then the elements of its leading diagonal cannot all be null.

18. (i) If A, B be symmetric matrices, then $A+B$ is also symmetic ; but AB is symmetric if, and only if, $AB = AB$.

(ii) The sum of two skew-symmetric matrices is skew-symmetric.

(ii) If Q be a skew-symmetric matrix, investigate the symmetry of Q^m, for $m = 0, 1, 2, \ldots$

19. Find an othogonal matrix having,

(i) $(\frac{1}{3}, \frac{2}{3}, -\frac{2}{3})$ as a row-vector ;

(ii) $(0, \frac{4}{5}, -\frac{3}{5})$ as a column-vector ;

(iii) $\left(\frac{1}{2}, -\frac{1}{2}, \frac{1}{\sqrt{2}}\right)$ and $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right)$ as row-vectors,

20. If A, B are square matrices, such that $\begin{pmatrix} A & C \\ O & B \end{pmatrix}$ is an orthogonal matrix, then A, B are orthogonal matrices and $C=O$.

21. If A, B are non-singular matrices such that $AA^t=BB^t$, then there exists an orthogonal matrix P such that $A=BP$.

22. Let S, Q, and R be respectively a symmetric, a skew-symmetric and an orthogonal matrix, then

(i) RSR^{-1} is symmetric,

(ii) RQR^{-1} is skew-symmetric, and

(iii) if $A=SR$, then $S^2=AA^t$.

4 Complex matrices.

Let the ground field be taken, in particular, to be the field of complex numbers C. A matrix over C will be called a *complex matrix*. A given complex matrix A can be expressed as $A=M+iN$, where M, N are real matrices (i.e., matrices over the field of real numbers). The matrix $A^*=M-i.N$ is called the *conjugate of the matrix* A. The elements of A^* are the conjugates of the corresponding elements of A. Evidently, $A^{**}=A$. Again if $B=P+i.Q$ be also a complex matrix, P, Q being real matrices, then

$$(AB)^*=[(M+i.N)(P+i.Q)]^*=[(MP-NQ)+i.(NP+MQ)]^*$$
$$=(MP-NQ)-i.(NP+MQ)=(M-i.N)(P-i.Q)=A^*B^*.$$

Also, $(A^t)^*=(M^t+iN^t)^*=M^t-i.N^t=(M-i.N)^t=(A^*)^t$.

Let $(A^t)^*=(A^*)^t$ be denoted by A°, then

$$(A^\circ)^\circ=([(A^*)^t]^t)^*=A.$$

Hermitian and skew-Hermitian matrices. A complex matrix H is said to be *Hermitian* if $H^\circ=H$, and *skew-Hermitian* if $H^\circ=-H$. Thus H is Hermitian or skew-Hermitian, according as

$$h_{ki}^*=h_{ik} \quad \text{or} \quad -h_{ik}, \quad \text{for } i,\ k=1, \quad , n;$$

and, in particular, a diagonal element $h_{kk}=h_{kk}^*$ or $-h_{kk}^*$.

Product of two Hermitian matrices is not, in general, Hermitian. In fact if A, B be two Hermitian matrices, then $(AB)^\circ=B^\circ A^\circ=BA$; hence AB is Hermitian if, and only if, $BA=AB$.

thus :

III.4.17 *If any set of Hermitian matrices from a group, then the group is abelian*

Let $H=M+i.N$, where the real part M and the imaginary part N are both real matrices, then $H^\circ=M^t-i.N^t$. Hence it follows that

III.4.18. *If $H=M+i.N$ be a Hermitian matrix, then*

(*i*) *the diagonal elements of H are all real numbers, and*

(*ii*) *the real part M is a symmetric and the imaginary part N is a skew-symmetric matrix.*

III.4.19. *If $S=M+i.N$ be a skew-Hermitian matrix, then*

(i) *the diagonal elements of S are all imaginary numbers, and*

(ii) *the real part M is a skew-symmetric and the imaginary part N is a symmetric matrix.*

In particular, a real Hermitian matrix $H=M$ is a real symmetric matrix and a real skew-Hermitian matrix $S=M$ is a skew-symmetric matrix.

Properties III.4.18 ii) and III.4.19(ii) are found to be useful in constructing Hermitian and skew-Hermitian matrices.

Unitary matrices. A complex matrix U is said to be *unitary* if $U.U^\circ=I$. Since U is a square matrix, the right inverse U° is the unique both-sided inverse U^{-1}, so that a unitary matrix is necessarily non-singular. Again,

$$U^\circ.U^{\circ\circ}=U^\circ.U=U^{-1}.U=I\,;$$

hence $U^{-1}=U^\circ$ is also unitary Also the product of two unitary matrices U, V is unitary, since

$$(UV)(UV)^\circ=(UV)[(UV)^t]^*=(UV).(V^tU^t)^*=(UV)[(V^t)^*(U^t)^*]$$
$$=(UV)(V^\circ U^\circ)=U(VV^\circ)U^\circ=UU^\circ=I.$$

Thus one obtains :

III.4.20. *The unitary matrices of order n form a sub-group $U(n,C)$ of the full linear group $GL'n,C)$.*

The vector space $V_n(C)$ is a unitary space, of dimension n, when the inner product of two vectors $\xi=(x_1,\ldots,x_n)$, $\eta=(y_1, \;, y_n)$ is defined by $\xi.\eta=x_1y_1^*+ \quad +x_ny_n^*$.

Then, from $U.U^\circ=I=U^\circ.U$, it follows, as in III.4.15 and III.4.16, that

III.4.21. *A complex matrix is unitary if, and only if, the row-vectors (column-vectors) form an orthonormal system.*

For a real matrix U, $U^{\circ}=U^{t}$; hence every real unitary matrix is an orthogonal matrix.

EXAMPLE 4 : Let $A=\begin{pmatrix} a+ib & -c+id \\ c+id & a-ib \end{pmatrix}$, then

$$A^{\circ}=\begin{pmatrix} a-ib & -c-id \\ c-id & a+ib \end{pmatrix}^{t}=\begin{pmatrix} a-ib & c-id \\ -c-id & a+ib \end{pmatrix}.$$

$$\therefore \quad A.A^{\circ}=\begin{pmatrix} a+ib & -c+id \\ c+id & a-ib \end{pmatrix}\begin{pmatrix} a-ib & c-id \\ -c-id & a+ib \end{pmatrix}$$

$$=\begin{pmatrix} a^2+b^2+c^2+d^2 & 0 \\ 0 & a^2+b^2+c^2+d^2 \end{pmatrix}.$$

Hence A is unitary if $a^2+b^2+c^2+d^2=1$.

Normal matrices. A complex matrix N is said to be a *normal matrix* , if $N.N^{\circ}=N^{\circ}N$.

All Hermitian, skew-Hermitiam and unitary matrices (including real symmetric, real skew-symmetric, real and complex orthogonal and diagonal matrices) are evidently normal. Besides these, there are others which are also normal :

EXAMPLE 5. **The following matrix P is normal, but it is neither Hermitian, nor skew-Hermitian, nor unitary :**

For $P=\begin{pmatrix} 0 & 2-i \\ 2-i & 0 \end{pmatrix}$, $P^{\circ}=\begin{pmatrix} 0 & 2+i \\ 2+i & 0 \end{pmatrix}$, and $P.P^{\circ}=\begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}=P^{\circ}.P.$

Since $(N^{\circ})^{\circ}=N$, it follows that, for a normal matrix N, the matrix N° is also normal ; $N^{-1}(=N^{\circ})$ is therefore normal. Product of two normal matrices is not normal (in general); the product is normal if, and only if they are commutative (the proof being similar to that of III.4.17).

Exercises

1. Every complex matrix can be expressed as a sum of a Hermitian and a skew-Hermition matrix.

2. Find which of the following are Hermitian or skew-Hermitian :

(i) $\begin{pmatrix} 2 & 3-2i \\ 3+2i & -1 \end{pmatrix}$, (ii) $\begin{pmatrix} -i & 2+3i \\ -2+3i & 2i \end{pmatrix}$,

(iii) $\begin{bmatrix} 2i & -1+i & 2-4i \\ 1+i & -i & -3 \\ -2-4i & 3 & 5i \end{bmatrix}$, (iv) $\begin{bmatrix} 1 & 2+2i & 4 \\ 2-2i & 2 & 2i \\ 4 & -2i & 3 \end{bmatrix}$.

3. Construct a Hermitian and a skew-Hermitian matrix by applying III. 4.18 and III. 4.19 on matrices :

$$M=\begin{bmatrix} -1 & 2 & 3 \\ 2 & 0 & -5 \\ 3 & -5 & 4 \end{bmatrix}, \quad N=\begin{bmatrix} 0 & 3 & -2 \\ -3 & 0 & 1 \\ 2 & -1 & 0 \end{bmatrix}.$$

4. Let H and U be respectively a Hermitian and a unitary matrix ; find the conditions so that the matrices

$$UHU^{-1}, \quad HUH^{-1}, \quad \begin{pmatrix} H & \\ & U \end{pmatrix}$$

may be (i) Hermitian, (ii) unitary.

5. The transposed and the conjugate of a unitary matrix are unitary.

6. Verify that the following matrices are unitary ;

(i) $\begin{pmatrix} \frac{1}{2}+\frac{1}{2}i & \frac{1}{2}-\frac{1}{2}i \\ \frac{1}{2}-\frac{1}{2}i & \frac{1}{2}+\frac{1}{2}i \end{pmatrix}$, (ii) $\begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}}+\frac{1}{\sqrt{3}}i \\ \frac{1}{\sqrt{3}}-\frac{1}{\sqrt{3}}i & -\frac{1}{\sqrt{3}} \end{bmatrix}$,

(iii) (a_{km}), where $a_{km}=\left[\cos\frac{2k(m-1)\pi}{n}+i.\sin\frac{2k(m-1)\pi}{n}\right]$, for $k=1, 2, \ldots, n$ and $m=1, 2, \ldots, n$.

7. Find a unitary matrix having $\left(\frac{1}{3}-\frac{2}{3}i, \; -\frac{1}{\sqrt{3}}+\frac{1}{3}i\right)$ as a row-vector.

8. The group $U(n, \mathbf{C})$ is isomorphic to a subgroup of $O(2n, E)$, where $\mathbf{C}$ and $\mathbf{E}$ are respectively the fields of complex and real numbers.

9. If a matrix has two of three properties in any one of the following three cases, it has all the three properties in that case :

(i) real, orthogonal, unitary ;
(ii) symmetric, orthogonal, involutory ;
(iii) Hermitian, unitary, involutory.

[A matrix A is said to be involutory if $A^2=I$.]

10. If N be a normal matrix, then, NU^{-1} is also normal, when U is any unitary matrix.

CHAPTER IV

DETERMINANTS

1. Determinant.

Let $A=(a_{ik})$ be a square matrix of order n over a field K. A certain function of the matrix whose value is an element of K, called the *determinant of the matrix* A, and denoted by

$$\det(A),\ |a_{ik}| \text{ or } \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

is determined by the following properties ;

$D(1)$: $\operatorname{Det}(I_m(c).A) = c \det(A)$,

$D(2)$: $\operatorname{Det}(I_{mk}(1).A) = \det(A)$,

$D(3)$: $\operatorname{Det}(I) = 1$.

Let $\alpha_k = (a_{k1}, a_{k2}, \ldots, a_{kn})$, for $k = 1, 2, \ldots, n$, be the row-vectors of the matrix A. Then $\det(A)$ may also be considered as a function of the row-vectors, possessing the following properties :

$D'(1)$: $\operatorname{Det}(\alpha_1, \ldots, \alpha_{m-1}, c\alpha_m, \alpha_{m+1}, \ldots, \alpha_n)$
$= c.\det(\alpha_1, \ldots, \alpha_n)$,

$D'(2)$: $\operatorname{Det}(\alpha_1, \ldots, \alpha_{m-1}, \alpha_m + \alpha_k, \alpha_{m+1}, \ldots, \alpha_n)$
$= \det(\alpha_1, \ldots, \alpha_n)$ for $k \neq m$,

$D'(3)$: $\operatorname{Det}(\varepsilon_1, \ldots, \varepsilon_n) = 1$, where $\varepsilon_r = (0, \cdots, 0, 1, 0, \ldots, 0)$ has 1 as the rth co-ordinate and 0 as the other co-ordinates.

The following propositions follow directly from the above properties :

IV.1.1. *If one of the row-vectors of A be null, then $\det(A) = 0$.*

Proof : Let $\alpha_i = \theta$, then $\alpha_i = 0.\alpha_i$, and

$$\begin{aligned} \det(A) &= \det(\alpha_1, \ldots, \alpha_{i-1}, \theta, \alpha_{i+1}, \ldots, \alpha_n) \\ &= \det(\alpha_1, \ldots, \alpha_{i-1}, 0\alpha_i, \alpha_{i+1}, \ldots, \alpha_n) \\ &= 0.\det(\alpha_1, \ldots, \alpha_{i-1}, \alpha_i, \alpha_{i+1}, \ldots, \alpha_n), \quad \text{by } D'(1) \\ &= 0. \end{aligned}$$

IV 1.2. *Det* $I_{ik}(c).A)=\det(A)$, *for* $i \neq k$.

Proof : The proposition is trivial when $c=0$. Next, let $c \neq 0$, then

$$\det(A = c^{-1}.\det(\alpha_1,\ldots,\alpha_i\ldots,\alpha_{k-1}, c\alpha_k,\alpha_{k+1}\ldots:\alpha_n)$$

$$= c^{-1}.\det(\alpha_{i1},\ldots,\alpha_{i-1}.\alpha_i+c\alpha_k,\alpha_{i+1},\ldots\alpha_{k-1}, c_k\alpha_k,\alpha_{k+1},\ldots,\alpha_n), \text{ by } D'(2)$$

$$= \det(\alpha_1,\ldots,\alpha_{i-1},\alpha_i+c\alpha_k,\alpha_{i+1}:\ldots,\alpha_{k-1},\alpha_k\,\alpha_{k+1},\ldots,\alpha_n), \text{ by } D'(1)$$

$$= \det(I_{ik}(c).A).$$

IV.1.3. *If the row-vectors of A be linearly dependent, then* $\det(A)=0$.

Proof : Late the row-vectors $\alpha_1,\ldots,\alpha_n$ of A be linearly dependent. Then some one of these, say α_k, can be expressed as a linear combination of the preceding vectors (II.1.11) :

$$\alpha_k = c_1\alpha_1+\ldots+c_{k-1}\alpha_{k-1}, \quad \text{or,} \quad \alpha_k - c_1\alpha_1 - \ldots - c_{k-1}\alpha_{k-1} = \theta.$$

Then the kth row-vector of the matrix, viz

$$I_{kk-1}(-c_{k-1})\ldots I_{k2}(-c_2)I_{k1}(-c_1)\,A$$

is the null vector θ, so that determinant of this matrix is 0. But determinant of this matrix is, by IV.1.2, equal to det (A) ; hence det $(A)=0$.

IV.1.4. *Det* $(I_{km}.A)=(-1).\det(A)$.

Proof : Det $(A)=\det(\alpha_1,\ldots\alpha_k,\ldots,\alpha_m,\ldots,\alpha_n)$

$$= \det(\alpha_1,\ldots,\alpha_k,\ldots,\alpha_{m-1},\alpha_m+\alpha_k,\alpha_{m+1},\ldots,\alpha_n)$$

$$= \det(\alpha_1,\ldots,\alpha_{k-1},\alpha_k+(-1)(\alpha_m+\alpha_k),\alpha_{k+1},\ldots,\alpha_{m-1},\alpha_m+\alpha_k,\alpha_{m+1},\ldots,\alpha_n)$$

$$= \det(\alpha_1,\ldots,\alpha_{k-1},-\alpha_m,\alpha_{k+1},\ldots,\alpha_{m-1},\alpha_m+\alpha_k,\alpha_{m+1},\ldots,\alpha_n)$$

$$= \det(\alpha_1,\ldots,\alpha_{k-1},-\alpha_m,\alpha_{k+1},\ldots,\alpha_{m-1},(\alpha_m+\alpha_k)+(-\alpha_m),\alpha_{m+1},\ldots,\alpha_n)$$

$$= \det(\alpha_1,\ldots,\alpha_{k-1},(-1)\alpha_m,\alpha_{k+1},\ldots,\alpha_{m-1},\alpha_k,\alpha_{m+1},\ldots,\alpha_n)$$

$$= (-1).\det(\alpha_1,\ldots,\alpha_{k-1},\alpha_m,\alpha_{k+1},\ldots,\alpha_{m-1},\alpha_k,\alpha_{m+1},\ldots,\alpha_n)$$

$$= (-1).\det(I_{km}.A).$$

Since a permutation is even or odd according as it is a product of an even or an odd number of transpositions, it follows from IV.1.4 that

IV.1.5. *A determinant remains unchanged by any even permutation of its row-vecors, whereas by an odd permutation it is transformed to its negative.*

Then, it follows from $D(3)$ that

IV.1.6. *Det* $(\varepsilon_{i_1},\dots,\varepsilon_{i_n})=1$ *or* -1 *according as the permutation* $\begin{pmatrix}1\dots n\\ i_1\dots i_n\end{pmatrix}$ *is even or odd, where* $\varepsilon_r=(0,\dots,0,1,0,\dots,0)$ *with* 1 *in the rth position.*

Now, it follows from IV.1.6, $D(1)$ and IV.1.2, by taking $A=I$, that :

IV.1.7. $Det(I_{km})=-1$, $det(I_m(c))=c$, $det\ (I_{km}(c))=1$.

An immediate consequence of this proposition and $D(1)$, IV.1.2 and IV.1.4. is that

IV.1.8. (i) *Det* $(EA)=det(E).\ det\ (A)$, *where* E *is an elementary matrix and* A *an arbitary matrix ;*

(ii) $det(D)=d_1d_2\dots d_n$. *where* $D=(d_i)$ *is a diagonal matrix.*

IV.1.9. $Det(DA)=det(D).det(A)$, *where* D *is a diagonal matrix and* A *an arbitrary matrix.*

Proof :
$$\begin{aligned}\text{Det}\ (DA)&=\det\ (I_1(d_1).\ I_2(d_2)\dots I_n(d_n)A)\\ &=d_1.\det\ (I_2(d_2)\dots I_n(d_n).A)\\ &=d_1d_2\dots d_n\ \det(A)=\det\ (D).\ \det\ (A).\end{aligned}$$

Corollary : *If* A *be a square matrix of order* n, *Det* $(-A)=$ $-1)^n.\ \det(A)$.

Proof : $-A=D.A$, where D is the diagonal matrix all whose diagonal elements are -1, so that $\det(D)=(-1)^n$.

The proposition IV.1 9. may now be generalised :

IV.1 10. $Det(AB)=det(A).det(B)$, *for any two matrices* A *and* B.

Proof : Let F_r denote the diagonal matrix having 1 in r diagonal positions, the other elements being 0.

The matrices A and B can be expressed (by III.4.3) as the products,

$$A=PF_rQ,\quad B=RF_m S,\quad \text{so that}\quad AB=PF_rQRF_mS,$$

where P, Q, R, S are products of elementary matrices.

If $r<n$, then $\det(F_r)=0$ (since some row-vectors of F_r are null); hence, it follows by IV.1.8(i) and IV.1.9 that

$$\det(A)=\det(P)\det(F_rQ)=\det(P)\det(F_r)\det(Q)=0,$$

and $$\det(AB)=\det(P)\det(F_r)\det(QRF_mS)=0.$$

Thus the proposition holds in this case. Similar is the case when $m<n$.

If $r=m=n$, one gets $F_r=F_m=I$; and then

$$\det(AB)=\det(PQRS=\det(PQ).\det(RS) \qquad \text{(by IV.1.8 (i))}$$
$$=\det(A).\det(B).$$

IV.1.11. *Let $A_i(\beta)$ be the matrix obtained by replacing the ith row-vector of A by the vector β, then*

$$\det(A_i(\beta+\gamma))=\det(A_i(\beta))+\det(A_i(\gamma)). \qquad \ldots(1)$$

Proof: If the row-vectors $\alpha_1,\ldots,\alpha_{i-1},\alpha_{i+1},\ldots,\alpha_n$ be linearly dependent, then (by IV.1.3) all the three determinants in (1) are null, and the proposition holds.

Next, let the row-vectors $\alpha_1,\ldots,\alpha_{i-1},\alpha_{i+1},\ldots,\alpha_n$ be linearly independent. Then two cases are possible:

(i) β may be linearly dependent on these, or

(ii) $\alpha_1,\ldots\alpha_{i-1},\beta,\alpha_{i+1},\ldots,\alpha_n$ are linearly independent.

In case (i), $\det(A_i(\beta))=0$ (by IV.1.3). Also, if

$$\beta=c_1\alpha_1+\ldots+c_{i-1}\alpha_{i-1}+c_{i+1}\alpha_{i+1}+\ldots+c_n\alpha_n,$$

then $I_{in}(-c_n)\ldots I_{ii+1}(-c_{i+1})I_{ii-1}(-c_{i-1})\ldots I_{i1}(-c_1).A_i(\beta+\gamma)=A_i(\gamma)$. Thus $\det(A_i(\beta+\gamma))=\det(A_i(\gamma))$ by IV.1.2, and the proposition holds.

In case (ii), the vectors $\alpha_1,\ldots,\alpha_{i-1},\beta,\alpha_{i+1},\alpha_n$ being linearly independent and n in number, form a basis of $V_n(K)$; hence γ can be expressed as a linear combination.

$$\gamma=b_1\alpha_1+\ldots+b_{i-1}\alpha_{i-1}+b_i\beta+b_{i+1}\alpha_{i+1}+\ldots+b_n\alpha_n.$$

Then replacing γ by $\gamma-b_1\alpha_1-\ldots-b_{i-1}\alpha_{i-1}-b_{i+1}\alpha_{i+1}-\ldots-b_n\alpha_n=b_i\beta$ in $\det(A_i(\gamma))$, one obtains

$$\det(A_i(\gamma))=\det(A_i(b_i\beta))=b_i.\det(A_i(\beta)).$$

Similarly, by replacing $\beta+\gamma$ by $\beta+\gamma-b_1\alpha_1-\ldots-b_{i-1}\alpha_{i-1}-b_{i+1}\alpha_{i+1}-\ldots-b_n\alpha_n=(1+b_i)\beta$, one obtains

$$\det(A_i(\beta+\gamma))=\det(A_i([1+b_i]\beta))=(1+b_i).\det(A_i(\beta))$$
$$=\det(A_i(\beta))+b_i.\det(A_i(\beta)$$
$$=\det(A_i(\beta))+\det(A_i(\gamma)).$$

IV.1.12. *A determinant is a linear function of each of its row-vectors.*

Proof: Let $\alpha_i = b_1\beta_1 + \ldots + b_r\beta_r$, then

$$\det(A) = \det(A_i(b_1\beta_1 + \ldots + b_r\beta_r))$$

$$= \det(A_i(b_1\beta_1)) + \det(A_i(b_2\beta_2 + \ldots + b_r\beta_r)) \text{ (by IV1.11)}$$

$$\det(A_i(b_1\beta_1)) + \det(A_i(b_2\beta_2)) + \ldots + \det(A_i(b_r\beta_r))$$

$$= b_1.\det(A_i(\beta_1)) + b_2.\det(A_i(\beta_2)) + \ldots + b_r.\det(A_i(\beta_r)).$$

Let ε_r be the vector having *1* as the *r*th co-ordinate and *0* as the other co-ordinates. Then the *i*th row-vector α_i of A can be expressed as

$$\alpha_i = (a_{i1}, \ldots, a_{in}) = a_{i1}\varepsilon_1 + \ldots + a_{in}\varepsilon_n, \text{ for } i = 1, 2, \ldots, n.$$

Hence, by IV.1.12, one obtains, for $i = 1, \ldots, n$,

IV.1.13. $\det(A) = a_{i1}\det(A_i(\varepsilon_1)) + \ldots + a_{in}\det(A_i(\varepsilon_n))$.

One can now express the determinant as a function of the elements :

IV.1.14. Existence theorem. *For a given matrix A, of order n, there exists a unique determinant function given by*

$$\det(A) = \Sigma(-1)^t a_{1r_1} a_{2r_2} \cdots a_{nr_n} \qquad \ldots (2)$$

where the summation extends over all permutation $r_1, \ldots, r_n$ of the suffices 1, 2,...,n, and t is an integer, even or odd, according as $\begin{pmatrix} 1 & 2 & \ldots n \\ r_1 & r_2 & \ldots r_n \end{pmatrix}$ is an even or an odd permutation.

Conversely, the right hand side expression, in (2) determines a determinant function of a square matrix A having $\alpha_i = (a_{i1}, \ldots, a_{in})$ as its i-th row-vector, for $i = 1, 2, \ldots, n$.

Proof: Taking $i = 1$ in IV.1.13, one gets

$$\det(A) = a_{11}\det A_1(\varepsilon_1) + \ldots + a_{1n}\det A_1(\varepsilon_n)$$

$$= \sum_{i=1}^{n} a_{1i} \det(\varepsilon_i, \alpha_2, \ldots, \alpha_n)$$

$$= \sum_{i=1}^{n} a_{1i} \det(\varepsilon_i, a_{21}\varepsilon_1 + \ldots + a_{2n}\varepsilon_n, \alpha_3, \ldots, \alpha_n)$$

$$= \sum_{i=1}^{n} a_{1i} \sum_{k=1}^{n} a_{2k} \det(\varepsilon_i, \varepsilon_k, \alpha_3, \ldots, \alpha_n)$$

$$\ldots \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots$$

$$= \sum_{i=1}^{n} a_{1i} \sum_{k=1}^{n} a_{2k} \cdots \sum_{m=1}^{n} a_{nm} \cdot \det(\varepsilon_i, \varepsilon_k, \ldots, \varepsilon_m)$$

$$= \Sigma a_{1i} a_{2k} \ldots a_{nm} \det(\varepsilon_i, \varepsilon_k, \ldots, \varepsilon_m),$$

where $i, k, \ldots, m$ assume all the values $1, 2, \ldots, n$ independently. But, by IV.1.6 and IV.1.3, $\det(\varepsilon_i, \varepsilon_k, \ldots, \varepsilon_m) = 1, -1$ or 0, according as $(i, k \ldots, m)$ is an even permutation of $1, 2, \ldots, n$, an odd permutation of $1, 2, \ldots, n$, or the suffices $i, k, \ldots, m$ are not different. Hence one obtains the expression (2) for $\det(A)$.

Conversely, let $\det(A)$ be defined to be the function of the elements of the matrix given by the summation expression (2), then it can readily be shown that $\det(A)$ possesses the three properties $D(1)$, $D(2)$, $D(3)$:

Each term of the sum contains just one element from each row as also from each column. Thus, if the ith row-vecor α_i be replaced by $c.\alpha_i$, then each term of the sum is multiplied by c ; consequently $\det(A)$ is changed to $c.\det(A)$. This proves $D(1)$.

To prove $D(2)$, IV.1.4 is proved first. If two of the second suffices, say i and k, be interchanged, the terms in the summation expression are permuted, and every odd permutation of the second suffices becomes even and conversely ; hence $\det(A)$ is transformed to $-\det(A)$. Whence it follows that if two row-vectors are equal then $\det(A) = 0$. If in A the row-vector α_m is replaced by $\alpha_m + \alpha_k$, the factor a_{mr} is replaced by $a_{kr} + a_{mr}$ in every term of the sum (2) and the term is therefore increased by $a_{1r_1} \cdots a_{kr_k} \cdot a_{nr_n}$. The sum of these additional terms taken with the corresponding sign + or − is equal to the determinant obtained from A by replacing α_m by α_k. Hence $\det(A)$ is altered to $\det(A) + \det(\alpha_1, \ldots, \alpha_k, \ldots, \alpha_{m-1}, \alpha_k, \alpha_{m+1}, \ldots, \alpha_n) = \det(A)$. This proves $D(2)$.

Finally, the only non-null term in the sum (2) corresponding to $\det(\varepsilon_1, \ldots, \varepsilon_n)$ is obtained by taking the two suffices equal in every factor ; consequently $\det(\varepsilon_1, \ldots, \varepsilon_n) = e_{11} e_{22} \ldots e_{nn} = 1$.

IV.1.15 $\quad$ *Det*$(A^t) = \det(A)$.

Proof : If the factors in any term, say $a_{1r_1} \ldots a_{nr_n}$, in the summation expression (2) be arranged in such a way that the second suffices are in the natural order, i.e., the term is arranged as $a_{s_1 1} \quad a_{s_n n}$ then the permutation

$\begin{pmatrix} 1 & 2 & \cdots & n \\ s_1 & s_2 & \cdots & s_n \end{pmatrix}$ is inverse of the permutation $\begin{pmatrix} 1 & 2 & \cdots & n \\ r_1 & r_2 & \cdots & r_n \end{pmatrix}$.

The two permutations are both odd or both even; hence the summation expression (2) behaves symmetrically with respect to the first and second indices. Now, a_{ik} for i fixed and $k=1, 2, \ldots, n$ are the co-ordinates of the row-vector α_i, and a_{ik} for k fixed and $i=1, 2, \ldots, n$ are the co-ordinates of the column-vector $\alpha^{(k)}$. Hence any function, symmetric in the two suffices i, k, remains unaltered when the row-vectors are replaced by the corresponding column-vectors, and vice versa. Consequently $\det(A^t)=\det(A)$.

The properties of a determinant function have been determined in terms of the row-vectors By the above proposition, corresponding properties hold in terms of column-vectors as well. Thus all the propositions of this article hold when column-vectors are considered in place of row-vectors. In particular, it follows from IV.1.13, that if $A^{(r)}(\varepsilon^s)$ be the matrix obtained from A by replacing the rth column-vector by the column-vector ε^s (having *1* in the sth column position and *0* elsewhere), then for $r=1, \; , n$,

$$\det(A)=a_{1r}\det(A^{(r)}(\varepsilon^1))+\ldots+a_{nr}\det A)^{(r)}(\varepsilon^n)).$$

It will be shown that $\det(A^{(r)}(\varepsilon^s))=\det(A_s(\varepsilon_r))$ (IV.1.21); hence:

IV.1.16. *Det*$(A)=a_{1i}$*det* $A_1(\varepsilon_i))+ \quad + a_{ni}$*det* $A_n(\varepsilon_i))$,
for $i=1, \; , n$

Determinants of orthogonal, Hermitian and unitary matrices may now be found out:

IV.1.17. *If P be an othogonal matrix, then* $[\det(P)]^2=1$.

Proof: $\mathrm{Det}(PP^t)=\det P).\det(P^t)=[\det(P)]^2$

Also, $\det(PP^t)=\det(I)=1$. Hence, $[\det(P)]^2=1$,

Let A be a matrix over the field of complex numbers. Since the conjugate of a sum is the sum of the conjugates, and the conjugate of a product is the product of the conjugates, it follows, by taking the conjugate of the relation (2) in IV.1.14. that $\det(A^*)=[\det[(A)]^*$. From this, one now readily obtains the following :

IV.1.18. *If H be a Hermitian matrix, then $\det(H)$ is a real number.*

Proof : $\mathrm{Det}(H)=\det(H^\circ)=\det(H^t)^*=[\det(H^t)]^*=[\det(H)]^*$. Hence $\det(H)$ is a real number.

IV.1.19. *For a unitary matrix U, $\det(U)$ has modulus 1.*

Proof : $\mathrm{Det}(U^\circ)=\det(U^t)^*=[\det(U^t)]^*=[\det(U)]^*$.

Hence, $\det(UU^\circ)=\det(U).\det(U^\circ)=\det(U).[\det(U)]^*$ $=|\det(U)|^2$.

But $\det(UU^\circ)=\det(I)=1$. Hence $|\det(U)|^2=1$, so that $|\det(U)|=1$.

Exercises

1. Show that the area of a parallelogram (or the volume of a parallelopiped), considered as a function of the vectors representing any set of co-terminal sides (or edges), satisfies the three properties of the determinant function. Hence obtain the expression for the area (or the volume) in the form of a determinant.

· If A be a matrix of order n, $\det(cA)=c^n.\det(A)$.

3. Determinant of a skew-symmetric matrix S of odd order is null, unless the characteristic of the ground field is 2.

[*Hints* : $\mathrm{Det}(S)=\det(-S^t)=(-1)^{2n-1}.\det(S^t)=-\det(S)$, hence $\det(S)=0$, where $2n-1$ is the order of S.]

4. Determinant of a skew-Hermitian matrix Q is a real or a pure imaginary number according as its order is even or odd.

[*Hints* : $\mathrm{Det}(Q)=(-1)^r[\det(Q)]^*$, where r is the order of Q]

5. Determinant of a triangular matrix (i.e., in the echelon form) is the product of its diagonal elements.

6. For a given matrix M, there exists a symmetric matrix S, such that $[\det(M)]^2 = \det(S)$.

7. For a given matrix M of even order, there exists a skew-symmetric matrix Q, such the $[\det(M)]^2 = \det(Q)$.

8. If N be a non-singular matrix, then

$$\det(N^{-1}) = [\det(N)]^{-1}.$$

Prove the following :

9. $$\begin{vmatrix} a & b & c \\ b & c & a \\ c & a & b \end{vmatrix} = (a+b+c)(a+\omega b+\omega^2 c)(a+\omega^2 b+\omega c),$$

where $\omega^3 = 1$ (and $\omega \neq 1$). Generalise.

The determinant is known as a *circulant.* The above circulant $= 3abc - a^3 - b^3 - c^3$; hence deduce

$$\begin{vmatrix} 2bc-a^2 & c^2 & b^2 \\ c^2 & 2ca-b^2 & a^2 \\ b^2 & a^2 & 2ab-c^2 \end{vmatrix} = (a^3+b^3+c^3-3abc)^2.$$

10. $$\begin{vmatrix} x & 0 & 0 & d \\ -1 & x & 0 & c \\ 0 & -1 & x & b \\ 0 & 0 & -1 & a \end{vmatrix} = ax^3+bx^2+cx+d.$$ Generalise.

11. $$\begin{vmatrix} x & a & b & 1 \\ \alpha & x & c & 1 \\ \alpha & \beta & x & 1 \\ \alpha & \beta & \gamma & 1 \end{vmatrix} = (x-\alpha)(x-\beta)(x-\gamma),$$ a, b, c being arbitrary. Generalise.

12. $$\begin{vmatrix} yz & x & x^2 \\ zx & y & y^2 \\ xy & z & z^2 \end{vmatrix} = \begin{vmatrix} 1 & x^2 & x^3 \\ 1 & y^2 & y^3 \\ 1 & z^2 & z^3 \end{vmatrix} = (x-y)(y-z)(z-x)(xy+yz+zx).$$

13. $$\begin{vmatrix} a^3 & a^2 & a & 1 \\ b^3 & b^2 & b & 1 \\ c^3 & c^2 & c & 1 \\ d^3 & d^2 & d & 1 \end{vmatrix} = (a-b)(a-c)(a-d)(b-c)(b-d)(c-d).$$ Generalise.

This determinant is called an *alternant* Show that the above alternant is equal to $-\begin{vmatrix} b^2c^2+a^2d^2 & bc+ad & 1 \\ c^2a^2+b^2d^2 & ca+bd & 1 \\ a^2b^2+c^2d^2 & ab+cd & 1 \end{vmatrix}.$

14. $$\begin{vmatrix} a & b & c & d \\ b & a & d & c \\ c & d & a & b \\ d & c & b & a \end{vmatrix} = -(a+b+c+d)(b+c-a-d)(c+a-b-d)(a+b-c-d).$$

15. $$\begin{vmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & a^2 & b^2 \\ 1 & c^2 & 0 & a^2 \\ 1 & b^2 & c^2 & 0 \end{vmatrix} = \begin{vmatrix} 0 & a & b & c \\ a & 0 & c & b \\ b & c & 0 & a \\ c & b & a & 0 \end{vmatrix} = a^4+b^4+c^4-2a^2b^2-2b^2c^2-2c^2a^2.$$

2. Cofactors and Minors.

Let $A=(a_{ik})$ be a matrix of order n over a field K. Then

$$\det(A_i(\varepsilon_k))=\det(\alpha_1,\ldots,\alpha_{i-1},\varepsilon_k,\alpha_{i+1},\ldots,\alpha_n)$$

is called the *cofactor* of the element a_{ik} in the matrix A, for $i,k=1,2,\ldots,n$. The cofactor $\det(A_i(\varepsilon_k))$ is usually denoted by A_{ik}. Then it follows from IV.1.13 and IV.1.16 that

$$\sum_{r=1}^{n} a_{ir}A_{ir}=a_{i1}A_{i1}+\ldots+a_{in}A_{in}=\det(A), \qquad \ldots \quad (1)$$

and $$\sum_{r=1}^{n} a_{ri}A_{ri}=a_{1i}A_{1i}+\ldots+a_{ni}A_{ni}=\det(A), \qquad \ldots \quad (2)$$

for $i=1,2,\ldots,n$.

Again, $\det(\alpha_1,\ldots,\alpha_i,\ldots,\alpha_{k-1},\alpha_i,\alpha_{k+1},\ldots,\alpha_n)=0$;
also $\det(\alpha_1,\ldots,\alpha_i,\ldots,\alpha_{k-1},\alpha_i,\alpha_{k+1},\ldots,\alpha_n)$

$$=\det(\alpha_1,\ldots,\alpha_i,\ldots,\alpha_{k-1},a_{i1}\varepsilon_1+\ldots+a_{in}\varepsilon_n,\alpha_{k+1},\ldots,\alpha_n)$$
$$=a_{i1}A_{k1}+a_{i2}A_{k2}+\ldots+a_{in}A_{kn}.$$

Thus, if $i\neq k$, $$\sum_{r=1}^{n} a_{ir}A_{kr}=a_{i1}A_{k1}+\ldots+a_{in}A_{kn}=0, \qquad \ldots \quad (3)$$

and similarly, $$\sum_{r=1}^{n} a_{ri}A_{rk}=a_{1i}A_{1k}+\ldots+a_{ni}A_{nk}=0, \qquad \ldots \quad (4)$$

for $i,k=1,2,\ldots,n$.

The results (1)—(4) may be stated in the form :

IV.2.1. (i) *If the elements of any row (or column) of a matrix A be multiplied by their respective cofactors and then added, the sum equals* $\det(A)$;

(ii) *if the elements of any row (or column) be multiplied by the cofactors of the corresponding elements of another row (resp. column) and then added, the sum is null.*

Cofactors are also obtained as determinants of submatrices of order $n-1$. In fact

IV.2.2. *The determinant of the matrix obtained by omitting the mth row and the rth column from the matrix A is equal to* $(-1)^{m+r}A_{mr}$. *(Consequently* $\det(A_m(\varepsilon_r))=A_{mr}$; *it is also equal to* $\det(A^{(r)}(\varepsilon_m))$*).*

Proof: The cofactor $A_{11}=\det(\varepsilon_1, \alpha_2,\dots, \alpha_n)$. Developing the determinant as a sum, as in (2), IV.1.14, one obtains

$$A_{11}=\Sigma(-1)^t e_{1s_1}\, a_{2s_2}\dots a_{ns_n},$$

where the summation extends over all permutations $s_1,\dots, s_n$ of the suffices $1, 2,\dots, n$, and t is an integer, even or odd according as $\begin{pmatrix}1 & 2 & \dots & n\\ s_1 & s_2 & \dots & s_n\end{pmatrix}$ is an even or an odd permutation.

Since $e_{1k}=0$ for $k\neq 1$ and $e_{11}=1$, it is sufficient to consider the value $s_1=1$ only; then the two permutations $\begin{pmatrix}1 & 2 & \dots & n\\ s_1 & s_2 & \dots & s_n\end{pmatrix}$ and $\begin{pmatrix}2 & \dots & n\\ s_2 & \dots & s_n\end{pmatrix}$ are the same. Consequently

$$A_{11}=\Sigma(-1)^t a_{2s_2}\dots a_{ns_n},$$

where the summation extends over all permutations $s_2,\dots,s_n$ of the suffices $2,\dots, n$, and t is an integer, even or odd, according as $\begin{pmatrix}2 & \dots & n\\ s_2 & \dots & s_n\end{pmatrix}$ is an even or an odd permutation. Then A_{11} is the determinant of the sub-matrix obtained by omitting the first row and the first column from the matrix A. Thus the proposition holds for $m=1$ and $r=1$.

Next, A_{mr} is the determinant of the matrix obtained by replacing the element a_{mr} by 1 and putting 0 in place of the other elements in the mth row. By interchanging the mth row successively with the $(m-1$ th, $(m-2)$th,..., 2nd and 1st rows, the mth row can be brought to the first place, the relative order of the remaining rows being not altered. By these $m-1$ interchanges of rows, A_{mr} is altered to $(-1)^{m-1}A_{mr}$

Now the rth column is similarly moved to the first position by $r-1$ interchanges of columns without altering the relative order of the other columns ; and let B be the resulting matrix. Then

$$\det(B)=(-1)^{m+r-2}A_{mr}=(-1)^{m+r}A_{mr}.$$

Now the first row-vector of B is $(1, 0, \ldots, 0)$, hence $\det(B) = B_{11}$ (by (1)). But, by what has been proved above, B_{11} is the determinant of the matrix obtained by omitting the first row and the first column from B. Hence B_{11} is also the determinant of the matrix obtained by striking out the mth row and the rth column from A, since the interchanges (of rows and columns) by which A has been transformed to B did not change the relative order of the rows and columns except for the mth row and the rth column (which have been brought respectively to the 1st row and the 1st column positions).

Adjoint or adjugate of a matrix. The matrix $B=(b_{ik})$, where $b_{ik}=A_{ki}$, for $i, k=1, 2, \ldots, n$, is called the *adjugate* or *adjoint* of the matrix A, and it is denoted by $adj(A)$. Then it follows immediately, by (1) — (4), that

IV.2.3. $A.adj(A)=adj(A).A=det(A).I.$

A matrix A is non-singular if, and only if, it is a product of elementary matrices (III.4.7) ; and since determinant of any elementary matrix is non-null (IV.1.7), it follows by IV.1.10 that

IV.2.4. *The determinant of a matrix is non-null if, and only if, it is non-singular.*

Now, one obtains readily from IV.2.3 that

IV.2.5. *A non-singular matrix A has the unique both sided inverse given by* $A^{-1}=\dfrac{1}{det(A)}adj(A)$.

EXAMPLE 1: Let P be an orthogonal matrix, then $P^t=P^{-1}$. Hence $p_{ik}=P_{ik}/\det(P)=\pm P_{ik}$ (by IV.1 17).

EXAMPLE 2: From the relation $A.\mathrm{adj}(A)=\det(A).I$ one obtains, by IV.1.10 and exercise 2, page 95,

$$\det(A).\det[\mathrm{adj.}(A)]=\det[\det(A).I]=[\det(A)]^n\ ;$$

thus

$$\det[\mathrm{adj}(A)]=[\det(A)]^{n-1}.$$

Minors. Let $B=(b_{ik})$ be an $m\times n$ matrix over a field K, and let C be a sub-matrix obtained by omitting the $j_1,\dots,j_{m-r}$th rows and $k_1,\dots,k_{n-r}$th columns from B. Then C is a square matrix of order r; $(-1)^t\det(C)$ is called a *minor* of order r of the matrix B, where

$$t=(j_1+\dots+j_{m-r})+(k_1+\dots+k_{n-r}).$$

For a square matrix A of order n, the cofactors of the elements are the minors of order $n-1$, with suitable signs. Since omission of a particular row and a particular column, say the mth row and the rth column, means replacing the mth row-vector by the fundamental unit vector ϵ_r in $(-1)^{m+r}\det(A)$, it follows that a minor of order r is obtained, with a suitable sign, by replacing $n-r$ row-vectors in $\det(A)$ by $n-r$ suitable fundamental unit vectors.

Let C be the minor of order $n-k$ obtained by omitting the $m_1,\dots,m_k$th rows and $r_1,\dots,r_k$th columns from $\det(A)$ (of order n), then C is obtained by replacing the $m_1,\dots,m_k$th row-vectors respectively by the fundamental unit vectors $\epsilon_{r_1},\dots,\epsilon_{r_k}$ in $(-1)^t\det(A)$, where $t=m_1+\dots+m_k+r_1+\dots+r_k$. The minor C will be denoted by $A_{m_1,\dots,m_k}(r_1,\dots,r_k)$. The minor of order k obtained by omitting all rows except the $m_1,\dots,m_k$th rows and all columns except the $r_1,\dots,r_k$th columns from $(-1)^t\det(A)$ is called the *complementary minor* of C.

Laplace's development of a determinant. A determinant has been expressed in (1) in terms of the elements of a row and their cofactors. Thus the determinant is expressed as a sum of products of minors of order 1, formed from a row, and their respective complementary minors. This has been generalised by Laplace to express the determinant as a sum of products of the minors of order m, formed from any m rows, and their respective complementary minors. For the purpose, the minors of order m, formed from the first m rows, of a matrix A of order n will be considered.

Every term in the expression

$$\det(A)=\Sigma(-1)^t a_{1r_1}\dots a_{nr_n} \quad\dots\quad (5)$$

can be split up into the two factors $(-1)^t(a_{1r_1}\dots a_{mr_m})$ and $(a_{m+1r_{m+1}}\dots a_{nr_n})$, where elements in the first factor belong to the first m rows (and those in the second belong to the remaining $n-m$ rows), and t is even or odd according as

$$\begin{pmatrix} 1\dots m & m+1\dots n \\ r_1\dots r_m & r_{m+1}\dots r_n \end{pmatrix} \qquad \dots \quad (6)$$

is an even or an odd permutation. Keeping $r_1,\dots,r_m$ the same, one can choose $r_{m+1},\dots,r_n$ such that the permutation (6) is even, so that the term may be initially taken as the product $(a_{1r_1}\dots a_{mr_m}).(a_{m+1r_{m+1}}\dots a_{nr_n})$. For any permutation $r'_{m+1},\dots,r'_n$ of the (second) suffices $r_{m+1},\dots,r_n$ in the second factor, the corresponding term is multiplied by $(-1)^q$, where q is even or odd according as the permutation

$$\begin{pmatrix} r_{m+1} & \dots r_n \\ r'_{m+1} & \dots r'_n \end{pmatrix} \qquad \dots \quad (7)$$

is even or odd, since the permutation

$$\begin{pmatrix} 1 & \dots & m & m+1 & \dots & n \\ r_1 & \dots & r_m & r'_{m+1} & \dots & r'_n \end{pmatrix}$$

is the product of the permutations (6) and (7), of which (6) is even. Hence all the terms in the sum (5), having $(a_{1r_1}\dots a_{mr_m})$ as a factor, collected together, with their respective signs, is $(a_{1r_1}\dots a_{mr_m}).A_{m+1,\dots,n}(r_{m+1},\dots,r_n)$. By the same argument, it follows that all the terms in the sum (5), having the minor $A_{m+1,\,,n}(r_{m+1},\,.\,,r_n)$ as a factor taken together, with their respective signs, equals $A_{1,\dots,m}(r_1,\dots,r_m).A_{m+1,\dots,n}(r_{m+1},\dots,r_n)$. Now taking all different permutations of the suffices $r_1,\dots,r_m$ out of $1,\dots,n$, one obtains:

IV. 2. 6. *Det*$(A)=\Sigma A_{1,\dots,m}(r_1,\dots,r_m).A_{m+1,\dots,n}(r_{m+1}\dots r_n)$, *the summation being taken for all suffices* $r_1,\dots,r_m,\dots,r_n$, *where* $\begin{pmatrix} 1 & \dots & m & m+1 & \dots & n \\ r_1 & \dots & r_m & r_{m+1} & \dots & r_n \end{pmatrix}$ *is an even permutation.*

Since any m rows can be brought into the first m positions, by a number of interchanges of rows, a similar result holds for the development of the determinant in terms of minors of order m, formed from any m rows of the given matrix.

EXAMPLE 3 : Let $B=(b_{ik})$ be a matrix of order 4. Then the minors of order 2, formed from the first two rows, are

$B_{12}(1,2)$, $B_{12}(1,3)$, $B_{12}(1,4)$, $B_{12}(2,3)$, $B_{12}(2,4)$ and $B_{12}(3,4)$.

Then $\det(B)=\Sigma B_{12}(p,q).B_{34}(r,s)$, where p, q, r, s, are so adjusted that the permutation $\begin{pmatrix} 1 & 2 & 3 & 4 \\ p & q & r & s \end{pmatrix}$ is even. So, when (p, q) equals (1, 2), (1, 3), (1, 4), (2, 3), (2, 4) or (3, 4), the corresponding pair (r, s) must be taken as (3, 4), (4. 2), (2, 3,) (1, 4), (3, 1) or (1,2) respectively. Thus Laplace's development is, in this case, given by

$$\det(B)=B_{12}(1,2).B_{34}(3,4)+B_{12}(1,3).B_{34}(4,2)+B_{12}(1,4).B_{34}(2,3)$$
$$+B_{12}(2,3).B_{34}(1,4)+B_{12}(2,4).B_{34}(3,1)+B_{12}(3,4).B_{34}(1,2).$$

An important result correlating minors of a given matrix and those of its adjoint matrix is given in the following proposition.

IV.2.7. Jacobi's theorem. *If M denotes any minor of order r of a matrix A of order n, and P denotes the corresponding minor of $adj(A)$, then the complementary minor N of P is given by*

$$N=M.[\det(A)]^{n-r-1}.$$

Proof : It is sufficient to prove the proposition when M denotes the minor of order r, obtained by omitting the last $n-r$ rows and $n-r$ columns. Let $\det(A)=\Delta$, then

$$\begin{bmatrix} a_{11} \ldots a_{1r} & a_{1r+1} \ldots a_{1n} \\ \ldots \quad \ldots & \ldots \quad \ldots \quad \ldots \\ a_{r1} \,.\, a_{rr} & a_{rr+1} \ldots a_{rn} \\ 0 \;\, .. \;\, 0 & 1 \quad .. \;\, 0 \\ \ldots \quad \ldots & \ldots \quad \ldots \quad \ldots \\ 0 \ldots 0 & 0 \;\ldots\; 1 \end{bmatrix} . \begin{bmatrix} A_{11} & .. \, A_{r1} & A_{r+11} & \ldots A_{n1} \\ \ldots & \ldots & \ldots & \ldots \\ A_{1r} & \ldots A_{rr} & A_{r+1r} & \ldots A_{nr} \\ A_{1r+1} & . \, A_{rr+1} & A_{r+1r+1} & A_{nr+1} \\ \ldots & \ldots & \ldots & \ldots \\ A_{1n} & \ldots A_{rn} & A_{r+1n} & \ldots A_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} \Delta & \ldots \; 0 & 0 & \ldots \; 0 \\ 0 & \ldots \; 0 & 0 & \ldots \; 0 \\ \ldots & \ldots & \ldots & . \;\; . \;\; \ldots \\ 0 & \Delta & 0 & \ldots \; 0 \\ A_{r+11} & \ldots A_{r+1r} & A_{r+1r+1} & \ldots A_{r+1n} \\ \ldots & \ldots & \ldots & \ldots \quad \ldots \\ A_{n1} & \ldots A_{nr} & A_{nr+1} & A_{nn} \end{bmatrix}.$$

Taking determinant of both sides of the above matrix equality, one gets $M.\det[\mathrm{adj}(A)]=\Delta^r.N$. Hence,

$$M.(\det(A))^{n-1}=(\det(A))^r.N, \quad \text{(by example 2, page 99)}$$

thus,

$$N=M[\det(A)]^{n-r-1}.$$

IV.2.8. *The determinant of a skew-symmetric matrix of even order is a perfect square.*

Proof: Let A be a skew-symmetric matrix of order $2m$, and let M be taken to be the principal minor of order $2m-2$, obtained by omitting the last two rows and the last two columns. Then M is the determinant of a skew-symmetric matrix of order $2m-2$.

$$\text{Here } N=\begin{vmatrix} 0 & A_{2m,\,2m-1} \\ A_{2m-1,\,2m} & 0 \end{vmatrix}=(A_{2m,\,2m-1})^2,$$

since $A_{2m-1,\,2m}=-A_{2m,\,2m-1}$. Hence

$$(A_{2m,\,2m-1})^2=M.[\det(A)]^{2m-(2m-2)-1}=M\det(A).$$

Thus the product of the determinants of two skew-symmetric matrices of orders $2m-2$ and $2m$ is a perfect square. Also the determinant of a skew-symmetric matrix of order 2, $\begin{vmatrix} 0 & a \\ -a & 0 \end{vmatrix}=a^2$, is a perfect square. Hence the result follows by applying the principle of mathematical induction.

Exercises

1. (*i*) Find the adjoint of the following matrices :

$$N=\begin{bmatrix} -1 & 3 & 4 \\ 2 & 0 & -1 \\ 1 & 2 & 3 \end{bmatrix}, \quad M=\begin{bmatrix} 1 & 2 & 3 \\ -2 & 1 & -1 \\ -1 & 3 & 2 \end{bmatrix},$$

$$S=\begin{bmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{bmatrix}, \quad T=\begin{bmatrix} a^2 & a & 1 \\ b^2 & b & 1 \\ c^2 & c & 1 \end{bmatrix}.$$

(*ii*) Find the adjoints of the elementary matrices I_{ij}, $I_r(c)$ and $I_{rs}(c)$.

2. For the matrices given in exercise 1, find $N^{-1}M$, MN^{-1}, T^{-1}, $T^{-1}ST$.

3. The adjoint of a skew-symmetric matrix of order n is symmetric or skew-symmetric according as n is odd or even.

4. (*i*) The adjoint and inverse of a Hermitian matrix are Hermitian.

(*ii*) The adjoint of a unitary matrix is unitary.

5. Let A, B be matrices of order n, then

(*i*) $\operatorname{adj}(AB)=\operatorname{adj}(A).\operatorname{adj}(B)$,

(*ii*) $A.\operatorname{adj}(B).B.\operatorname{adj}(A)=\det(AB).I$,

(*iii*) $\operatorname{adj}[\operatorname{adj}(A)]=[\det(A)]^{n-2}A$,

(*iv*) $\det(\operatorname{adj}[\operatorname{adj}(A)])=[\det(A)]^m$, where $m=(n-1)^2$.

6. Matrices having determinant *1* or *−1* form a group, called the *extended unimodular group*; it is a normal subgroup of the full linear group $GL(n, K)$.

The *unimodular group*, consisting of all matrices with determinant *1*, is also a normal subgroup of $GL(n, K)$.

7. If P, Q be orthogonal matrices such that $\det(P)=-\det(Q)$, then $\det(P+Q)=0$.

8. Let $\beta_i=(b_{i1},\dots, b_{in})$, $i=1, \dots, n$, be the row-vectors of a matrix B, over the field of real numbers, then

$$|\det(B)| \leqslant \|\beta_1\|\dots\|\beta_n\|,$$

where the equality holds if $\beta_1, \dots, \beta_n$ form an orthogonal system.

If moreover $-g \leqslant b_{ik} \leqslant g$, for $i, k=1,\dots, n$, then

$$\det(B) \leqslant (g\sqrt{n})^n. \quad [\textit{Hadamard's determinant theorem}]$$

[*Hints*: $BB^t=(\beta_i.\beta_k)$ and $\|\beta_i\|\leqslant g\sqrt{n}$.]

9. In an orthogonal matrix of order 4, every minor of order 2 is equal to its complementary minor.

10. If A be an $m\times n$ matrix and B an $n\times m$ matrix, where $m\leqslant n$, then $\det(AB)$ is the sum of products of minors of order m of A and the *corresponding* minors of order m of B.

[*Binet-Cauchy theorem*]

11. If P be a real $m\times n$ matrix, then $\det(PP^t)>0$.

[For $m=2$, the result reduces to *Cauchy-Schwarz inequality*.]

12. If A and B are square matrices, then

$$\det\begin{pmatrix} A & C \\ 0 & B \end{pmatrix}=\det(A).\det(B).$$

13. A matrix, whose diagonal elements are null, is called a *zero-axial matrix*. Any matrix can be developed in terms of the diagonal elements and zero-axial determinants:

$$\begin{vmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ a_4 & b_4 & c_4 & d_4 \end{vmatrix} = \begin{vmatrix} 0 & b_1 & c_1 & d_1 \\ a_2 & 0 & c_2 & d_2 \\ a_3 & b_3 & 0 & d_3 \\ a_4 & b_4 & c_4 & 0 \end{vmatrix} + a_1 \begin{vmatrix} 0 & c_2 & d_2 \\ b_3 & 0 & d_3 \\ b_4 & c_4 & 0 \end{vmatrix}$$

$$+ b_2 \begin{vmatrix} 0 & c_1 & d_1 \\ a_3 & 0 & d_3 \\ a_4 & c_4 & 0 \end{vmatrix} + c_3 \begin{vmatrix} 0 & b_1 & d_1 \\ a_2 & 0 & d_2 \\ a_4 & b_4 & 0 \end{vmatrix} + d_4 \begin{vmatrix} 0 & b_1 & c_1 \\ a_2 & 0 & c_2 \\ a_3 & b_3 & 0 \end{vmatrix}$$

$$+ a_1 b_2 \begin{vmatrix} 0 & d_3 \\ c_4 & 0 \end{vmatrix} + a_1 c_3 \begin{vmatrix} 0 & d_2 \\ b_4 & 0 \end{vmatrix} + a_1 d_4 \begin{vmatrix} 0 & c_2 \\ b_3 & 0 \end{vmatrix}$$

$$+ b_2 c_3 \begin{vmatrix} 0 & d_1 \\ a_4 & 0 \end{vmatrix} + b_2 d_4 \begin{vmatrix} 0 & c_1 \\ a_3 & 0 \end{vmatrix} + c_3 d_4 \begin{vmatrix} 0 & b_1 \\ a_2 & 0 \end{vmatrix} + a_1 b_2 c_2 d_4.$$

14. Show that

$$\begin{vmatrix} 0 & a & b & c \\ -a & 0 & d & e \\ -b & -d & 0 & f \\ -c & -e & -f & 0 \end{vmatrix} = (af - be + cd)^2.$$

15. *The skew determinant,*

$$\begin{vmatrix} x & a & b & c \\ -a & x & d & e \\ -b & -d & x & f \\ -c & -e & -f & x \end{vmatrix} = x^4 + (a^2 + b^2 + c^2 + d^2 + e^2 + f^2)x^2 + (af - be + cd)^2.$$

[*Hints* : Apply excrises 13 and 14].

16. Let $f(x) = (x-\alpha)(x-\beta)(x-\gamma)(x-\delta)$ over the field of real numbers, then

$$\begin{vmatrix} \alpha & x & x & x \\ y & \beta & x & x \\ y & y & \gamma & x \\ y & y & y & \delta \end{vmatrix} = \frac{yf(x) - xf(y)}{y - x}.$$

For $y = x$, the determinant $= f(x) - xf'(x)$.

17. If $H = A + iB$ be Hermitian, where A is non-singular, then

$$[\det(H)]^2 = [\det(A)]^2 . \det(I + (A^{-1}B)^2).$$

3. Minors and rank.

It follows from (1), page 97 that if every minor of order $n-1$ of a matrix A be null, then $\det(A)=0$. Since cofactors of the elements of a submatrix of order $n-1$ are minors of order $n-2$ (of A), it follows that if every minor of order $n-2$ be null, the same holds for all minors of order $n-1$ and hence also for $\det(A)$. Repeating this argument, one obtains :

IV.3.1. *If every minor of order r of a matrix A be null, then all minors of higher order are also null.*

IV.3 2. *If a minor of order r of an $m\times n$ matrix A be non-null whereas all minors of higher order are null, then $\text{rank}(A)=r$.*

Proof : Without any loss of generality, one may assume that $m \leqslant n$. Since the rank of a matrix is not altered by any permutation of its row or column-vectors, one can assume, without any loss of generality, that the minor, formed by the elements common to the first r rows and the first r columns, is non-null, i.e., $\det(B)\neq 0$, where

$$B=\begin{pmatrix} a_{11} & \dots & a_{1r} \\ \dots & \dots & \dots \\ a_{r1} & \dots & a_{rr} \end{pmatrix}.$$

Let
$$F=\begin{pmatrix} a_{11} & \dots & a_{1r} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{r1} & \dots & a_{rr} & \dots & a_{rn} \end{pmatrix}$$

be the matrix formed by the row-vectors $\alpha_1,\dots,\alpha_r$ of A, then
$$\text{rank}(F)=\text{rank}[R(F)] \leqslant r.$$
Also
$$\text{rank}(F)=\text{rank}[C(F)] \geqslant r,$$
since the first r column-vectors of F, being the column-vectors of B, are linearly independent (since det $B\neq 0$, i.e., B is non-singular). Thus rank $(F)=r$; consequently the row-vectors $\alpha_1,\dots,\alpha_r$ are linearly independent

Consider the matrices G_u, for a particular value of k, where $r<k\leqslant m$, and

$$G_u=\begin{pmatrix} a_{11} & \dots & a_{1r} & a_{1u} \\ \dots & \dots & \dots & \dots \\ a_{r1} & \dots & a_{rr} & a_{ru} \\ a_{k1} & \dots & a_{kr} & a_{ku} \end{pmatrix}$$

for $u=1,\ldots,n$. The cofactors of the elements in the last column of G_u remain the same for all values of u (since these depend on the first r columns of G_u); and let these be denoted by $T_1,\ldots, T_r, T_k$ respectively, then (by (2), page 97):

$$a_{1u}T_1+\ldots+a_{ru}T_r+a_{ku}T_k=\det(G_u).$$

For $u=1,\ldots, r$, $\det(G_u)=0$, since two column-vectors are identical. Also, for $u=r+1,\ldots, n$, $\det(G_u)$ is a minor of order $r+1$ of the matrix A; hence $\det(G_u)=0$, by hypothesis. Thus

$$a_{1u}T_1+\ldots+a_{ru}T_r+a_{ku}T_k=0,$$

for $u=1,\ldots, n$; consequently

$$T_1\alpha_1+\ldots+T_r\alpha_r+T_k\alpha_k=\theta.$$

Again

$$T_k=A_{ku}=\det(B)\neq 0.$$

Hence α_k is linearly dependent on $\alpha_1,\ldots, \alpha_r$, (which are linearly independent) for all values of k, where $r<k\leqslant m$. Hence

$$\operatorname{rank}(A)=\operatorname{rank}[R(A)]=r.$$

Example 1: If A be a matrix of order n and rank r, then rank [adj (A)] $=n$, 1 or 0, according as $r=n, n-1$ or $<n-1$.

(I) Let $r=n$, then A is non-singular; and therefore $\det(A)\neq 0$. Then $\det[\operatorname{adj}(A)]=[\det(A)]^{n-1}$ (vide example 2, page 99) $\neq 0$. Hence adj(A) is non-singular, and therefore rank [adj (A)] $=n$.

(II) Let $r<n$, then $\det(A)=0$. In the notation of Jacobi's theorem, order of $N=n-t$. For $n-t>1$ (i.e., when $t<n-1$), $N=M.[\det(A)]^{n-t-1}=0$. Thus, when $\det(A)=0$, every minor of order >1 of adj(A) is null. In other words,

$$\text{if } \det(A)=0, \text{ then rank adj}(A)\leqslant 1. \qquad \ldots \qquad (1)$$

(i) Let $r=n-1$. Then there exists at least one minor N of order $n-1$ of A, such that $M=0$. That is, the cofactor of one at least of the elements of A is non-null. Thus, one at least of the elements of adj(A) is non-null. Hence rank adj$(A)\geqslant 1$. Combining this with (1), one obtains the result, viz. rank adj$(A)=1$.

(ii) Let $r<n-1$. Then every minor of order $n-1$ of A is null. That is, cofactor of every element of A is null. Thus every element of adj(A) is null. Hence rank adj$(A)=0$.

Exercises

1. Find the ranks of the matrices given in exercise 1, page 66, and exercise 2, page 72, by the method of minors.

2. Show that all the diagonal elements of a symmetric or a skew-symmetric matrix of rank 1 cannot be 0. Hence, deduce that there does not exist any skew-symmetric matrix of rank 1.

3. A symmetric matrix A is of rank r if, and only if,

   (i) there exists a non-null principal minor S_r of order r of A, and

   (ii) all principal minors of orders $r+1$ and $r+2$, containing S_r, are null

The same is also true for a Hermitian matrix A.

4. Defining the rank of a matrix as the order of a non-null minor of maximum order, show that the rank of a matrix remains unaltered on pre- or post-multiplcation by an elementary matrix.

5. If A be an $m \times n$ matrix of rank r, and B and C are non-singular matrices of orders m and n respectively, then the product matrices BA and AC have rank r.

6. Let A be a square matrix of order n, over a field K. Then det $(xI-A)$ is called the characteristic polynomial of the matrix A. Show that the coeffieient of $(-1)^k x^{n-k}$ in the characteristic polynomial of A is the sum of all principal minors of order k of A.

CHAPTER V

SOLUTION OF SYSTEM OF LINEAR EQUATIONS

Let K be a given field, consider a system of m linear equations in n unknowns $x_1, \ldots, x_n$,

$$\left.\begin{array}{l} a_{11}x_1 + \ldots + a_{1n}x_n = a_{1n+1} \\ \ldots \quad \ldots \quad \ldots \quad \ldots \\ a_{m1}x_1 + \ldots + a_{mn}x_n = a_{mn+1} \end{array}\right\} . \qquad \ldots (1)$$

When $a_{1n+1}, \ldots, a_{mn+1}$ are all null, the system is called a *homogeneous system*, otherwise a *non-homogeneous system.*

1. Homogeneous system.

For a homogeneous system, the equations (1) reduce to

$$\left.\begin{array}{l} a_{11}x_1 + \ldots + a_{1n}x_n = 0 \\ \ldots \quad \ldots \quad \ldots \\ a_{m1}x_1 + \ldots + a_{mn}x_n = 0 \end{array}\right\} . \qquad \ldots (2)$$

The n-tuples, $\alpha_i = (a_{i1}, \ldots, a_{in})$, for $i = 1, \ldots, m$, are called the *coefficient vectors*, and the matrix.

$$A = \begin{bmatrix} a_{11}, \ldots, a_{1n} \\ \ldots \quad \ldots \quad \ldots \\ a_{m1}, \ldots, a_{mn} \end{bmatrix} \qquad \ldots (3)$$

is called the *cofficient matrix* of the system (2).

Let $V_n(K)$ be the vector space consisting of all n-tuples with co-ordinates from K, then $\operatorname{rank}[V_n(K)] = n$ (example 1, page 31). The vectors $\alpha_1, \ldots, \alpha_m$ belong to $V_n(K)$. The sub-space, generated by $\alpha_1, \ldots, \alpha_m$, is called the *coefficient space* C of the system (2).

An n-tuple $\xi = (x_1, \ldots, x_n)$ is a solution of the system (2) if $\alpha_i . \xi = 0$, for $i = 1, \ldots, m$. Thus ξ is orthogonal to every coefficient vector α_i, and hence to every vector of C (by II.3.12), and conversely a vector ξ, orthogonal to every vector of C, is a solution. Hence, by II.3.17, it follows that

V.1.1. *The solutions of the homogeneous system (2) form a sub-space X, which is the orthogonal complement of C in $V_n(K)$.*

The sub-space X is called the *solution space* of the system (2) ; it is also called the *null-space* of the sub-space C or that of that of the matrix A.

Since $V_n(K)$ is the direct sum of C and X,

$$V_n(K)=C+X, \text{ and } C\cap X=\theta.$$

Hence, $\text{rank}[V_n(K)]=\text{rank}(C)+\text{rank}(X)$ (by II.2 7), since $\text{rank}\,(C\cap X)=0$. Again $\text{rank}[V_n(K)]=n$, and $\text{rank}(C)=\text{rank}(A)$. Thus one obtains :

V.1.2. *Rank*$(X)=n-$*rank*$(C)=n-$*rank*(A).

Rank (X) is called the *nullity* of the sub-space C. Thus $\text{rank}(X)=\text{nullity of } C=n-\text{rank}(C)=n-\text{rank}\,(A)=\text{nullity}\,(A)$.

The null-tuple $(0, \ldots, 0)$ is always a solution of the system (2) ; *there exist non-null solutions if, and only if, rank*$(X)>0$, *i. e., rank* $(A)<n$, *i.e., det*$(A)=0$.

Method of solution. To obtain the solution space X, an orthonormal (orthogonal) basis of the coefficient space C, say $\gamma_1, \ldots, \gamma_r$, is found out. This set is then extended to an orthonormal (orthogonal) basis, $\gamma_1, \ldots, \gamma_r, \gamma_{r+1}, \ldots, \gamma_n$, of $V_n(K)$ (by II.3.15). Then the vectors $\gamma_{r+1}, \ldots, \gamma_n$ form a basis of the solution space X. This process is usually laborious. A simpler method is given below :

Method of sweep-out. The matrix A can be reduced, by suitable elementary row-operations to a fully swept out matrix (III 2.2) :

$$B=\begin{bmatrix} 1 & 0 & \ldots & 0 & -b_{1r+1} & \ldots & -b_{1n} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 & -b_{rr+1} & \ldots & -b_{rn} \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 0 & 0 & & 0 \end{bmatrix} \quad \ldots \quad (4)$$

(or a matrix which differs from the above by a permutation of the columns only). The non-null row vectors $\beta_1, \ldots, \beta_r$ of the matrix B form a basis of the coefficient space C. Hence a vector $\xi=(x_1, \ldots, x_n)$ is a solution of the system (2) if, and only if, it is orthogonal to the vectors $\beta_1, \ldots, \beta_r$, so that $x_1, \ldots, x_n$ satisfy the following equations :

$$\left.\begin{array}{l} x_1 - b_{1r+1}x_{r+1} - \ldots - b_{1n}x_n = 0 \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\ x_r - b_{rr+1}x_{r+1} - \ldots - b_{rn}x_n = 0 \end{array}\right\},$$

or,

$$\left.\begin{array}{l} x_1 = b_{1r+1}x_{r+1} + \ldots + b_{1n}x_n \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\ x_r = b_{rr+1}x_{r+1} + \ldots + b_{rn}x_n \end{array}\right\}.$$

A solution is then obtained by choosing any values for $x_{r+1}, \ldots, x_n$, and adjusting $x_1, \ldots, x_r$ as above. Hence the solutions of the system (2) are all given by

$$\xi = (b_{1r+1}t_{r+1} + \ldots + b_{1n}t_n, \ldots, b_{rr+1}t_{r+1} + \ldots + b_{rn}t_n, \; t_{r+1}, \ldots, t_n), \qquad \ldots (5)$$

where the parameters $t_{r+1}, \ldots, t_n$ run over K independently. Taking $(t_{r+1}, \ldots, t_n)$ equal to the fundamental unit vectors $(1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$, the following particular solutions are obtained :

$$\left.\begin{array}{l} \xi_1 \;\;\;\; = b_{1r+1}, \ldots, b_{rr+1}, 1, 0, \ldots 0) \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\ \xi_{n-r} = (b_{1n}, \ldots, b_{rn}, \;\; 0, 0, \ldots, 1) \end{array}\right\}. \qquad \ldots (6)$$

These vectors generate the solution space X, since the general solution (5) can be expressed as a linear combination of these,

$$\xi = t_{r+1}\xi_1 + \ldots + t_n\xi_{n-r}. \qquad \ldots (7)$$

The vectors (6) are also linearly independent, since a linear combination (7) of the vectors can be the null-vector if, and only if, $t_{r+1} = \ldots = t_n = 0$ (these being the last $n-r$ coordinates of ξ in (5)). Hence $\xi_1, \ldots, \xi_{n-r}$ form a basis of the solution space X.

EXAMPLE 1 : Consider the following homogeneous system :

$$\begin{array}{r} 2x_1 + 8x_2 + 4x_3 + \;\; x_4 = 0 \\ 4x_1 + 5x_2 + 5x_3 + 11x_4 = 0 \\ x_1 + \;\, x_2 + \tfrac{1}{2}x_3 + \;\, 5x_4 = 0 \\ 6x_1 + 7x_2 + 6x_3 + 21x_4 = 0. \end{array}$$

The coefficient matrix A is reduced to the fully swept out form B by some row operations :

$$A = \begin{bmatrix} 2 & 3 & 4 & 1 \\ 4 & 5 & 5 & 11 \\ 1 & 1 & \tfrac{1}{2} & 5 \\ 6 & 7 & 6 & 21 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & -\tfrac{5}{2} & 14 \\ 0 & 1 & 3 & -9 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = B.$$

A solution vector $\xi=(x_1, x_2, x_3, x_4)$ must be orthogonal to the vectors $\beta_1=(1, 0, -\frac{5}{2}, 14)$ and $\beta_2=(0, 1, 3, -9,)$, so that

$$\left.\begin{aligned} x_1-\tfrac{5}{2}x_3+14x_4&=0\\ x_2+3x_3-9x_4&=0 \end{aligned}\right\}, \qquad \text{i.e.,}\quad \left.\begin{aligned} x_1&=\tfrac{5}{2}x_3-14x_4\\ x_2&=-3x_3+9x_4 \end{aligned}\right\}.$$

Hence the solutions are all given by

$$\xi=(\tfrac{5}{2}u-14v, -3u+9v, u, v),$$

where u, v, are parameters. A basis of the solution space is given by the two vectors (corresponding to the values $u=1$, $v=0$ and $u=0$, $v=1$):

$$\xi_1=(\tfrac{5}{2}, -3, 1, 0). \quad \xi_2=(-14, 9, 0, 1).$$

Exercises

1. Solve the following systems of equations:

(i)
$$\begin{aligned} x-3y+2z&=0\\ -2x+6y+3z&=0\\ 5x-13y+10z&=0. \end{aligned}$$

(ii)
$$\begin{aligned} 2x-y+5z+3w&=0\\ x+y+z&=0\\ 5y-2z+7w&=0. \end{aligned}$$

(iii)
$$\begin{aligned} 2x-y-3z&=0\\ x+y+z&=0\\ 3x+6y+8z&=0. \end{aligned}$$

(iv)
$$\begin{aligned} -x+3y+2z+5w&=0\\ 2x-y+3z+w&=0\\ 7x-y+14z+9w&=0. \end{aligned}$$

2. Show that the following two systems have the same solutions:

$$\begin{aligned} 4x+3y-5z&=0\\ -x+y+3z&=0\\ 3x+2y-2z&=0, \end{aligned} \qquad \begin{aligned} 2x+3y+z&=0\\ x+6y+4z&=0\\ x+y-3z&=0. \end{aligned}$$

2. Non-homogeneous system.

Let the system (1) be non-homogeneous, then $(a_{1n+1},\ldots, a_{mn+1})\neq(0,\ldots,0)$.

The solutions of a non-homogeneous system (1) *do not form a sub-space*, since the null vector is not a solution. Let $(t_1,\ldots,t_n)$ be a solution of the system (1), then

$$\left.\begin{aligned} a_{11}t_1+\ldots+a_{1n}t_n&=a_{1n+1}\\ \ldots\quad \ldots\quad &\ldots\quad \ldots\\ a_{m1}t_1+\ldots+a_{mn}t_n&=a_{mn+1} \end{aligned}\right\} \qquad \ldots (3)$$

Subtracting the equalities (8) from the corresponding ones in (1), one finds

$$\left.\begin{array}{l} a_{11}(x_1-t_1)+\ldots+a_{1n}(x_n-t_n)=0 \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\ a_{m1}(x_1-t_1)+\ldots+a_{mn}(a_n-t_n)=0 \end{array}\right\} .$$

Thus $(x_1-t_1,\ldots,x_n-t_n)$ is a solution of the homogeneous system (2). Hence,

V.2.1. *The solutions of the system* (1) *are obtained by adding any particular solution of the system* (1) *to all vectors of the sub-space formed by the solutions of the homogeneous system* (2) [*obtained from* (1) *by putting* $a_{1n+1}=\ldots=a_{mn+1}=0$].

But a non-homogeneons system may *not* have any solution at all. E.g. the system

$$\left.\begin{array}{l} 2x_1-\ 5x_2=7 \\ 4x_1-10x_2=9 \end{array}\right\}$$

has no solution. Thus it is necessary to consider the question of solvability of a given non-homogeneous system.

V.2.2. *A system of linear non-homogeneous equations* (1) *is solvable if, and only if, rank* $(A')=$ *rank* (A), *where*

$$A=\begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots \\ a_{m1} & \ldots & a_{mn} \end{bmatrix}, \qquad A'=\begin{bmatrix} a_{11} & \ldots & a_{1n} & a_{1n+1} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & \ldots & a_{mn} & a_{mn+1} \end{bmatrix}.$$

Proof: Let the system (1) be solvable, then there exist elements, not all null, $x_1,\ldots,x_n \varepsilon K$ such that

$$\begin{bmatrix} a_{11} \\ \ldots \\ a_{m1} \end{bmatrix} x_1+\ldots+\begin{bmatrix} a_{1n} \\ \ldots \\ a_{mn} \end{bmatrix} x_n=\begin{bmatrix} a_{1n+1} \\ \ldots \\ a_{mn+1} \end{bmatrix},$$

Then the last column-vector of A' is linearly dependent on the first n column-vectors. Hence $\text{rank}[C(A')]=\text{rank}\,[C(A)]$; consequently, $\text{rank}(A')=\text{rank}(A)$.

Converse follows by reversing the above argument.

EXAMPLE 1: *If* $m=n$ *and* $\det(A)\neq 0$, *then the equations* (1), *page* 109, *have a unique solution.* In fact, $\det(A)\neq 0$ implies $\text{rank}(A)=n$; and then $\text{rank}(A')=n$. Thus the system (1) is solvable. Also, since $\text{rank}(A)=n$, the equations (2) have the null-vector as the only solution. Hence the result follows by V.2 1.

Method of solution. The solutions of a given system (1), when solvable, may be obtained as follows :

The given system (1) is transformed, by the substitution

$$x_1=y_1/y_{n+1},\ldots, x_n=y_n/y_{n+1}, \qquad \ldots \quad (9)$$

to a homogeneous system

$$\left.\begin{array}{l} a_{11}y_1+\ldots+a_{1n}y_n-a_{1n+1}y_{n+1}=0 \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\ a_{m1}y_1+\ldots+a_{mn}y_n-a_{mn+1}y_{n+1}=0 \end{array}\right\} . \qquad \ldots \quad (10)$$

If $\eta=(y_1,\ldots,y_n, y_{n+1})$ be the general solution of the system (10), then $\xi=(x_1,\ldots, x_n)$ is the general solution of the given system (1) where $x_1,\ldots, x_n$ are obtained by (9).

The solutions of (10) can generally be divided into two classes, (i) solutions for which $y_{n+1} \neq 0$, and (ii) solutions for which $y_{n+1}=0$. Solutions of the system (1) are obtained from the solutions of class (i). In the non-solvable case, every solution of the system (10) belongs to class (ii).

EXAMPLE 2 : To find the value of k for which the following system is solvable, and then to solve it :

$$\left.\begin{array}{r} x+\ y+\ z=2 \\ 2x+\ y+3z=1 \\ x+3y+2z=5 \\ 3x-2y+\ z=k \end{array}\right\} . \qquad \ldots \quad (11)$$

The matrix A' is swept out, by row-operations only, to

$$A'=\begin{bmatrix} 1 & 1 & 1 & -2 \\ 2 & 1 & 3 & -1 \\ 1 & 3 & 2 & -5 \\ 3 & -2 & 1 & -k \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -k-2 \end{bmatrix} = B'.$$

Then the swept out form of the matrix A is obtained by omitting the last column of B', i.e.,

$$A=\begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & -2 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = B.$$

Thus rank$(A)=3$. In order that the given system be solvable, rank (A') must also be 3, consequently $k=-2$.

To find the solutions, when $k=-2$, the given system is transformed, by the subsitution $x=y_1/y_4$, $y=y_2/y_4$, $z=y_3/y_4$, to the homogeneous system :

$$\left.\begin{aligned} y_1+y_2+y_3-2y_4&=0\\ 2y_1+y_2+3y_3-y_4&=0\\ y_1+3y_2+2y_3-5y_4&=0\\ 3y_2-2y_2+y_3+2y_4&=0 \end{aligned}\right\}. \qquad \ldots \quad (12)$$

The fully swept out form of the coefficient matrix is

$$\begin{bmatrix} 1 & 0 & 0 & -1\\ 0 & 1 & 0 & -2\\ 0 & 0 & 1 & 1\\ 0 & 0 & 0 & 0 \end{bmatrix}$$

A solution vector (y_1, y_2, y_3, y_4) of (12) is orthogonal to each of the row vectors of the above matrix ; hence,

$$y_1-y_4=0, \quad y_2-2y_4=0, \quad y_3+y_4=0.$$

So, the general solution of the transformed system (12) is given by $(y_1, y_2, y_3, y_4)=(u, 2u, -u, u)$, u being a parameter. Hence the general solution of the given system (11) is

$$(x, y, z)=(u/u, 2u/u, -u/u)=(1, 2, -1).$$

Thus (1, 2, −1) is the only solution of the given system.

Exercises

1. Solve, if possible, the following systems of equations :

178 (i) $5x+4y-8z=9$
$2x-3y+6z=2$
$7x+2y-4z=1.$

(ii) $x+y+z=5$
$2x-y+3z=7$
$x+3y+2z=2$
$x-4y+z=5$
$2x+y+4z=4.$

(iii) $7x-38y+3z=13$
$3x+13y+2z=17$
$2x-5y+z=6.$

(iv) $x-y+z-w=2$
$2x+3y-5z+w=1$
$x-6y-8z-4w=5.$

(v) $x+(2)y-z=(3)$
$(2)x-y+(3)z=(0).$

The equations (v) have coefficients from $GF(5)$.

2. Find the value of k for which the following systems of equations are solvable, and then solve them :

(i) $3x+4y-3z=5$
$x+2y-z=7$
$4x+y+2z=9.$
$2x-y+4z=k.$

(ii) $2x+y-z=3$
$x-y+3z=5$
$5x+y+z=11$
$3x+4y+2z=k.$

3. Consider the following system of equations, where a_i, b_i, c_i, d_i are rational numbers :

$$\begin{aligned} a_1x+a_2y+a_3z&=1\\ b_1x+b_2y+b_3z&=\sqrt{2}\\ c_1x+c_2y+c_3z&=\sqrt{3}\\ x+d_2y+d_3z&=0. \end{aligned}$$

Is it possible that for a suitable system of coefficients,

(i) the system of the four equations is solvable ;

(ii) the system of the first three equations is solvable ;

(iii) the system of the first three equations has more than one solution ?

3. Intersection of planes in Euclidean 3-space E^3.

Let a system of planes in E^3 be given by

$$\left.\begin{aligned} a_1x+b_1y+c_1z+d_1&=0\\ \cdots\quad\cdots\quad\cdots&\quad\cdots\\ a_kx+b_ky+c_kz+d_k&=0 \end{aligned}\right\} \qquad \ldots (1)$$

By the substitution $x=x_1/x_4$, $y=x_2/x_4$, $z=x_3/x_4$, the equations (1) are transformed to

$$\left.\begin{aligned} a_1x_1+b_1x_2+c_1x_3+d_1x_4&=0\\ \cdots\quad\cdots\quad\cdots\quad\cdots&\quad\cdots\\ a_kx_1+b_kx_2+c_kx_3+d_kx_4&=0 \end{aligned}\right\} \qquad \ldots (2)$$

which are the equations of the given planes, each extended by a line at infinity, in the extended Cartesian 3-space. Let

$$A=\begin{bmatrix} a_1 & b_1 & c_1\\ \cdots & & \cdots\\ a_k & b_k & c_k \end{bmatrix}. \qquad A'=\begin{bmatrix} a_1 & b_1 & c_1 & d_1\\ \cdots & & \cdots & \cdot\;\cdot\\ a_k & b_k & c_k & d_k \end{bmatrix},$$

then $\text{rank}(A')=\text{rank}(A)$ or $\text{rank}(A)+1$, and $\text{rank}(A)\leqslant 3$, $\text{rank}(A')\leqslant 4$. The nature of intersection of the given planes in E^3 is ascertained from the values of $\text{rank}(A')$ and $\text{rank}(A)$ as follows.

(I) *Rank*$(A')=4$. The rank of the solution space of the system (2) is 0. The planes have no common point in the extended Cartesian space, and can therefore have no common point in E^3.

(II) *Rank*$(A')=3$. The rank of the solution space of the system (2) is 1. The planes intersect at a point in the extended

Cartesian space ; the point is either a point in E^3, or it is a point at infinity.

Rank$(A)=3$ or 2. (i) If *rank*$(A)=3$, the system (1) is solvable ; hence the planes intersect in a point in E^3. (ii) If *rank*$(A)=2$, the system (1) is not solvable ; hence the given planes have no common point in E^3, but intersect at a point at infinity. Thus the lines of intersection of pairs of planes are parallel (so the planes are not all parallel or coaxial but are all parallel to a particular line).

(III) *Rank*$(A')=2$. The rank of the solution space of the system (2) is 2. The planes intersect in a line in the extended Cartesian space ; this line is either a line in E^3 with one point at infinity on it, or it is a line at infinity.

Rank$(A)=2$ or 1. (i) If *rank*$(A)=2$, the system (1) is solvable ; hence the planes intersect in a line in E^3, i.e., the planes are coaxial. (ii) If *rank*$(A)=1$, the system (1) is not solvable ; hence the planes intersect in a line at infinity, so that the planes are parallel.

(IV) *Rank*$(A')=1$. The rank of the solution space of the system (2) is 3. The planes are coincident.

The results may be put in the following tabular form :

Rank(A')	*Rank*(A)	*Arrangement of the planes.*
4	3	Planes have no common point,
3	3	Planes intersect in one point.
	2	Planes form a prism (planes are parallel to a line but are not themselves *all* parallel or coaxial).
2	2	Planes are coaxial.
	1	Planes are parallel.
1	1	Planes are coincident.

Exercises

1. Find the equation of the plane parallel to the plane $x-2y+3z=0$ and passing through the intersection of the planes,

$$\begin{aligned} x+8y-11z&=-11\\ 3x-\;y+\;2z&=\;\;12\\ 2x+\;y-\;\;z&=\;\;\;5. \end{aligned}$$

2. Determine the arrangement of the planes in Euclidean 3-space in the following systems :

$$\text{(i)}\quad \begin{aligned} -x+\;y+2z&=2\\ 2x-\;y+3z&=4\\ 3x+2y-6z&=1\\ 7x-3y-\;z&=5. \end{aligned} \qquad\qquad \text{(ii)}\quad \begin{aligned} 2x+3y+\;4z&=1\\ 5x+8y+13z&=1\\ x+\;y-\;\;z&=2\\ 3x+5y+\;9z&=1. \end{aligned}$$

4. Method of determinants.

Let a system of linear equations, which is solvable, be given by (1), page 109. *If* $m \leqslant n$, *and if a minor of order* m *of the coefficient matrix* A *be non-null, i e., the row-vectors of the coefficient matrix be linearly independent*, then the following method may be applied to solve the given system,

Without any loss of generality, one can assume that the minor Δ formed by the first m column-vectors of the coefficient matrix is non-null. The given system can be represented, in matrix notation, as follows :

$$\begin{pmatrix} a_{11} & a_{12}\ldots a_{1m} \\ a_{21} & a_{22}\ldots a_{2m} \\ \ldots & \ldots\;\ldots\;\ldots \\ a_{m1} & a_{m2}\ldots a_{mm} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_m \end{pmatrix} + \begin{pmatrix} a_{1m+1} \\ a_{2m+1} \\ \ldots \\ a_{mm+1} \end{pmatrix} x_{m+1} +$$

$$\ldots + \begin{pmatrix} a_{1n} \\ a_{2n} \\ \ldots \\ a_{mn} \end{pmatrix} x_n = \begin{pmatrix} a_{1n+1} \\ a_{2n+1} \\ \ldots \\ a_{mn+1} \end{pmatrix}.$$

Multiplying by the adjoint of the $m \times m$ matrix (a_{ik}), where $1 \leqslant i \leqslant m$, $1 \leqslant k \leqslant m$, one obtains

$$\Delta I.\begin{pmatrix}x_1\\x_2\\\cdots\\x_m\end{pmatrix}+\begin{pmatrix}A_{11}&A_{21}\cdots&A_{m1}\\A_{12}&A_{22}&A_{m2}\\\cdots&\cdots\;\cdot\cdot&\cdot\cdot\\A_{1m}&A_{2m}\cdot\cdot&A_{mm}\end{pmatrix}\left\{\begin{pmatrix}a_{1m+1}\\a_{2m+1}\\\cdots\\a_{mm+1}\end{pmatrix}x_{m+1}+\cdots+\begin{pmatrix}a_{1n}\\a_{2n}\\\cdots\\a_{mn}\end{pmatrix}x_n-\begin{pmatrix}a_{1n+1}\\a_{2n+1}\\\cdots\\a_{mn+1}\end{pmatrix}\right\}=\begin{pmatrix}0\\0\\\cdots\\0\end{pmatrix},$$

or,

$$\Delta\begin{pmatrix}x_1\\\cdots\\x_m\end{pmatrix}+\begin{pmatrix}\Delta^{(1)}(m+1)\\\cdots\\\Delta^{(m)}(m+1)\end{pmatrix}x_{m+1}+\cdots+\begin{pmatrix}\Delta^{(1)}(n)\\\cdots\\\Delta^{(m)}(n)\end{pmatrix}x_n-\begin{pmatrix}\Delta^{(1)}(n+1)\\\cdots\\\Delta^{(m)}(n+1)\end{pmatrix}=\begin{pmatrix}0\\\cdots\\0\end{pmatrix};$$

thus

$$\Delta x_r=-\Delta^{(r)}(m+1).x_{m+1}-\cdots-\Delta^{(r)}(n).x_n+\Delta^{(r)}(n+1),\quad\ldots(1)$$

for $r=1,\ldots,m$, where $\Delta^{(r)}(k)$ is obtained from Δ by replacing the rth column by the kth column of the coefficient matrix A' (for $k=m+1,\ldots,n+1$).

The solutions of the system (1), page 109, are given by (1) above, where $x_{m+1},\ldots,x_n$ are free to assume any values from K.

The method is applicable to non-homogeneous as also to homogeneous systems.

In the non-homogeneous case, when $n=m$, the solution (1) above can be expressed in the form ;

$$\frac{x_1}{\Delta^{(1)}(m+1)}=\cdots=\frac{x_m}{\Delta^{(m)}(m+1)}=\frac{1}{\Delta};\qquad\ldots(2)$$

this is known as *Crammer's rule.*

EXAMPLE 1 : **For a given non-homogeneous system, viz.**

$$\begin{aligned}x+2y+5z-4w&=6\\2x-3y-z+5w&=4\\4x+y-3z+2w&=8,\end{aligned}$$

$$\Delta=\begin{vmatrix}1&2&5\\2&-3&-1\\4&1&-3\end{vmatrix}=84,\qquad\Delta^{(1)}(4)=\begin{vmatrix}-4&2&5\\5&-3&-1\\2&1&-3\end{vmatrix}=41.$$

$$\Delta^{(1)}(5)=\begin{vmatrix} 6 & 2 & 5 \\ 4 & -3 & -1 \\ 3 & 1 & -3 \end{vmatrix}=143.$$

Hence $84x=-41w+143$, $\therefore$ $x=(-41w+143)/84$.
Likewise, $y=(101w-47)/84$ and $z=(85w+91)/84$.

Exercises

1. Apply the method of determinants to solve, where possible, the exercises 1 and 2, page 115.

2. Solve, by the method of determinants, the following systems of equations :

(i)
$$\begin{aligned} 2x-3y+4z&=3 \\ x+5y-7z&=-1 \\ -3x+7y-2z&=2, \end{aligned}$$

(ii)
$$\begin{aligned} x-2y+z+3w&=-3 \\ 2x+y+3z-5w&=1 \\ 4x-3y+5z+2w&=8, \end{aligned}$$

(iii)
$$\begin{aligned} 3x+y-4z+2w-t&=1 \\ x-2y+5z-w+2t&=5 \\ 5x-3y+7z-5w-3t&=1. \end{aligned}$$

3. Apply the method of determinants to solve the following system of equations over a field K :

$$\begin{aligned} a_1x+b_1y+c_1z+d_1w&=0 \\ a_2x+b_2y+c_2z+d_2w&=0 \\ a_2x+b_3y+c_2z+d_3w&=0. \end{aligned}$$

CHAPTER VI

EQUIVALENCE OF MATRICES

1. Equivalence of matrices over a field

Let K be a given field, and Ω the set of all $m \times n$ matrices over K. A matrix B is said to be *equivalent* to a matrix A, where $A, B \varepsilon \Omega$, if there exist non-singular matrices P, Q (of suitable sizes) such that $B=PAQ$. The relation of being equivalent is certainly an equivalence relation. In fact, (i) $A=IAI$, (ii) $B=PAQ$ implies $A=P^{-1}BQ^{-1}$, and (iii) $B=PAQ$ and $C=MBN$ imply $C=(MP)A(QN)$ prove the reflexive, symmetric and transitive properties since (i) the identity matrices are non-singular, (ii) the inverses of non-singular matrices are non-singular, and (iii) the products of non-singular matrices are non-singular. Accordingly, Ω is partitioned into disjoint classes of equivalent matrices.

Again, since a non-singular matrix is a product of some elementary matrices, and multiplication by elementary matrices mean elementary row or column operations, it follows that the equivalence of matrices as defined above is the same as that defined in terms of elementary row and column operations in page 65, Hence it follows from III.2.3 that :

VI.1.1. (i) *Any $m \times n$ matrix A over a field K is equivalent to an $m \times n$ matrix F in the fully reduced normal form*

$$F=\begin{pmatrix} I_r & O \\ O & O \end{pmatrix},$$

where the sub-matrix I_r is the identity matrix of order r, where r is the rank of A, and all other elements in F are null.

A complete set of canonical matrices is obtained by taking $r=0,1,2,\ldots, m$ or n (which-ever is less).

(ii) *Two $m \times n$ matrices over a field K are equivalent if and only if, they have the same rank.*

Exercises

1. Reduce the following matrices to their fully reduced normal forms, under equivalence :

$$A=\begin{pmatrix} 2 & -5 & 12 & 7 \\ 15 & 3 & -8 & 0 \\ -3 & 18 & 10 & 6 \end{pmatrix}, \qquad B=\begin{pmatrix} -5 & 3 & 1 & 14 \\ 2 & 10 & -4 & 6 \\ 12 & -7 & 15 & -9 \end{pmatrix}.$$

Are the matrices A and B equivalent ?

2. Find the non-singular matrices P_1, Q_1 and P_2, Q_2, such that P_1AQ_1 and P_2BQ_2 are in the fully reduced normal forms, where A, B are the two matrices given in exercise 1.

2. Equivalence of bilinear forms.

Let K be a given field. An expression,

$$\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}x_iy_j, \qquad \text{where } a_{ij} \varepsilon K, \qquad \dots (1)$$

is called a *bilinear form*, over K, in the two sets of variables $x_1, \dots, x_m$ and $y_1, \dots, y_n$. It is called *bilinear*, as it is linear in each of the two sets of variables. And, as it is a homogeneous polynomial in each set of variables, it is called a *form*.

The bilinear form (1) can be expressed as a matrix product XAY^t, where $X=(x_1, \dots, x_m)$, $Y=(y_1, \dots, y_n)$ are 1-row matrices (and Y^t is therefore a 1-column matrix), and $A=(a_{ij})$ is an $m \times n$ matrix over K. A is called the *matrix of the bilinear form*, and rank (A) is called the *rank of the bilinear form*.

A bilinear form XBY^t is said to be *equivalent* to a bilinear form XAY^t, if the matrix B is equivalent to the matrix A, over K. That is, $B=PAQ$, where P and Q are non-singular matrices (of suitable sizes) over K. In this case, XAY^t is transformed to $X'BY'^t$ by the substitutions $X=X'P$ and $Y=Y'Q$. Conversely, a given form XAY^t is equivalent to any form $X'BY'^t$ to which it is transformed by substitutions $X=X'P$ and $Y=Y'Q$, where P, Q are non-singular matrices.

It now follows directly from III.2.3. and VI.1.1(i) that

VI.2.1. *Any given bilinear form of rank r is equivalent to the canonical normal form* $x_1y_1+\dots+x_ry_r$.

It also follows directly from VI.1.1(ii) that

VI.2.2. *Two bilinear forms, with $m\times n$ matrices over a field K, are equivalent if, and only if, they have the same rank.*

Exercises

1. Reduce the following bilinear forms to their canonical normal forms, and find the the matrices of the substitutions by which the reductions are effected :

(i) $2x_1y_1-5x_1y_2+4x_1y_3-x_2y_1+3x_2y_2+7x_2y_3$;

(ii) $3x_1y_1-x_1y_2+x_1y_3-7x_1y_4+2x_2y_1-x_2y_2+5x_2y_3$
$-6x_2y_4+x_3y_1+2x_3y_2-5x_3y_3+9x_3y_4$.

3. Equivalence of matrices over a Euclidean domain.

Let Λ_n be the set of all square matrices of order n over a Euclidean domain W. A matrix B is said to be *equivalent* to a matrix A, where $A, B \varepsilon \Lambda_n$, if $B=PAQ$, P, Q being non-singular matrices over W, whose inverses are also matrices over W. A non-singular matrix is a product of elementary matrices (vide III.4.7.). For elementary matrices I_{rs} and $I_{rs}(c)$ of types I and III (vide page 74), over W, their inverses I_{rs} and $I_{rs}(-c)$ are also matrices over W. The same is true of an elementary matrix $I_r(c)$ of type II, over W, if, and only if, c^{-1} also belongs to W, *i.e.*, c is a unit in W. Thus P Q must be products of arbitrary elementary matrices of types I and III, and of those elementary matrices $I_r(c)$, of type II, for which c is a unit in W. Hence a matrix B, which is equivalent to a given matrix A, over a Euclidean domain W, is obtained from A by applying elementary row and column operations of types I and III (vide page 62), and by multiplication of a row or a column by any unit of W.

As in the case of equivalence over a field, the relation of being equivalent, over a Eucliden domain, is also an equivalence relation. One obtains, accordingly a classification of the matrices in A_n. The normal forms are given by :

VI.3.1. *Any square matrix A, of order n, over a Euclidean domain, is equivalent to a diagonal matrix in the Smith's normal form :*

$$S = \begin{bmatrix} e_1 & & & & & & \\ & \cdot & & & & & \\ & & \cdot & & & & \\ & & & e_r & & & \\ & & & & 0 & & \\ & & & & & \cdot & \\ & & & & & & \cdot \\ & & & & & & & 0 \end{bmatrix}, \quad \dots \quad (1)$$

in which e_i is a divisor of e_{i+1}, for $i=1,\dots r-1$, where r is the rank of the matrix A.

Proof: Let $A=(a_{ik})$ be a given square matrix of order n over a Euclidean domain W. The theorem is obvious when rank of A is 0. Next, let rank $(A)>0$. Among the non-null elements in A, there is one whose norm is least. It may be brought to the (1,1)-position by interchanges of rows and columns. So, it may be assumed, without any loss of generality, that the norm of a_{11} is the least. If $a_{m1} \neq 0$, for $m>1$, then $a_{m1}=qa_{11}+b$, where either $b=0$, or $b\mathrm{N}<a_{11}\mathrm{N}$. Hence adding $(-q)$ times the first row to the mth row, one obtains b in the $(m,1)$-position ; and if $b \neq 0$, it is brought to the (1,1)-position by an interchange of rows. Thus in a finite number of steps the given matrix is reduced to one in which all the elements in the first column, except that in (1,1)-position, are made null. Similarly, applying suitable column operations, all the elements in the first row, except that in the (1,1)-position, are made null. Let $B=(b_{ik})$ be the transformed matrix, then

$$B=\begin{pmatrix} b_{11} & O \\ O & B_1 \end{pmatrix}.$$

If now b_{11} be not a divisor of every non-null element in the matrix B_1, then b_{11} can be replaced (by suitabale row and column operations) by an element whose norm is less than $b_{11}N$. In fact, if b_{ik} be not divisible by b_{11}, then $b_{ik}=pb_{11}+c$, where $c \neq 0$ and $cN<b_{11}N$. Now, by adding the ith row to the first row, and then adding $(-p)$ times the first column to the kth column, one obtains c in the $(1,k)$-position; it can then be brought to the (1,1)-positon by interchange of columns. In this way the matrix B may be transformed to one, say

$$C=\begin{pmatrix} e_1 & 0 \\ 0 & C_1 \end{pmatrix},$$

where e_1 is a divisor of every non-null element in C_1. The element e_1 is evidently the h.c.f. of the elements in C; and therefore it is also the h.c.f. of the elements in A (since the h.c.f. of the elements in A remain unaltered by row and column operations that have been applied).

Next, by applying suitable row and column operations, without affecting the first row and the first column of C, the sub-matrix C_1 may be reduced to a matrix

$$D_1=\begin{pmatrix} e_2 & 0 \\ 0 & D_2 \end{pmatrix},$$

where e_2 is the h.c.f. of the elements in D_1 and also of those in C_1. Hence e_1 is a divisor of e_2, and A is equivalent to the matrix

$$D=\begin{pmatrix} e_1 & & \\ & e_2 & \\ & & D_2 \end{pmatrix}.$$

By continuing this process, one obtains, in a finite number of steps, the diagonal matrix S with the required properties.

EXAMPLE 1: To find the Smith's normal form of the matrix :

$$A=\begin{pmatrix} 6 & 4 & 8 & 10 \\ 4 & 12 & 20 & 24 \\ 14 & 20 & 16 & 18 \\ 8 & 12 & 22 & 16 \end{pmatrix}.$$

The h.c.f. of the elements in A is 2 ; it is obtained in the (1,1) position by adding (−1)-times the second row to the first row. And then by sweeping out the first row and the first column, one obtains :

$$A\to\begin{pmatrix} 2 & -8 & -12 & -14 \\ 4 & 12 & 20 & 24 \\ 14 & 20 & 16 & 18 \\ 8 & 12 & 22 & 16 \end{pmatrix}\to\begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 28 & 44 & 52 \\ 0 & 76 & 100 & 116 \\ 0 & 44 & 70 & 72 \end{pmatrix}=\begin{pmatrix} 2 & \\ & C_1 \end{pmatrix}$$

The h.c.f. of the elements of the sub-matrix C_1 is 2. To get it in the leading position, (i) (−1)-times the second-column is added to the third column, (ii) first and third rows are interchanged, and then (iii) the first and third columns are interchanged :

$$C_1=\begin{pmatrix} 28 & 44 & 52 \\ 76 & 100 & 116 \\ 44 & 70 & 72 \end{pmatrix} \to \begin{pmatrix} 28 & 44 & 8 \\ 76 & 100 & 16 \\ 44 & 70 & 2 \end{pmatrix}$$

$$\to \begin{pmatrix} 44 & 70 & 2 \\ 76 & 100 & 16 \\ 28 & 44 & 8 \end{pmatrix} \to \begin{pmatrix} 2 & 70 & 44 \\ 16 & 100 & 76 \\ 8 & 44 & 28 \end{pmatrix}$$

(and now sweeping out the first column and the first row)

$$\to \begin{pmatrix} 2 & 0 & 0 \\ 0 & -460 & -276 \\ 0 & -286 & -148 \end{pmatrix} = \begin{pmatrix} 2 & \\ & D_2 \end{pmatrix}.$$

The h.c.f. of the elements of the sub-matrix D_2 is 4. Now 4 is also the h.c.f. of −460 and −286. The division algorithms by which 4 may be obtained from −460 and −236 are

$$-460+(-2)(-286)=12\,; \quad -236+20.12=4.$$

Applying corresponding row-operations, one obtains

$$D_2=\begin{pmatrix} -460 & -276 \\ -286 & -148 \end{pmatrix} \to \begin{pmatrix} 12 & 20 \\ -236 & -148 \end{pmatrix}$$

$$\to \begin{pmatrix} 12 & 20 \\ 4 & 252 \end{pmatrix} \to \begin{pmatrix} 4 & 252 \\ 12 & 20 \end{pmatrix} \to \begin{pmatrix} 4 & 0 \\ 0 & -786 \end{pmatrix},$$

by (i) adding (−2)-times the second row to the first row, (ii) adding 20-times the first row to the second row, (iii) interchanging the rows, and then (iv) by sweep out.

Hence the required Smith's normal form is given by

$$\begin{pmatrix} 2 & & & \\ & 2 & & \\ & & 4 & \\ & & & -736 \end{pmatrix}.$$

To obtain the canonical forms, some invariants of a given matrix A, under equivalence, will now be studied.

Determinant divisors. Let $d_k(A)$ denote the h.c.f. of the set of all minors of order k of A, for $k=1,2,\ldots,r$, where r is the rank of A ; then $d_k(A)$ is called the kth *determinant divisor* of A.

Invariance of the determinant divisors of a matrix, under equivalence, is proved in the following theorem.

VI.3.2. *If B is equivalent to a matrix A, of rank r, over a Eulidean domain W, then $d_m(B)=d_m(A)$, for $m=1,2, \ldots, r$.*

Proof: It is sufficient to prove the theorem for a matrix B which has been obtained from A by applying any one of the row operations I and III (page 62) or by multiplying any row by a unit in W. If $B=I_{ik}A$, minors of the mth order of B are the same as those of A, except possibly for their signs. The same is also true if any row of A has been multiplied by a unit in W. Next, let $B=I_{ik}(c).A$. Then a minor of B in which the ith row takes no part or in which both ith and kth rows take part is the same as the corresponding minor of A. Also any minor of B, in which, ith row takes part but kth row does not, is equal to (the corresponding minor of A)+(c-times another minor of the same order). Hence, $d_m(A)$ is a divisor of $d_m(B)$. Again, since the relation of equivalence between A and B is symmetric, it follows that $d_m(B)$ is also a divisor of $d_m(A)$. Hence $d_m(A)$ is an associate of $d_m(B)$ in W, and one can take $d_m(A)=d_m(B)$, for $m=1,2, \ldots r$.

Invariant factors. Next, let $d_0(A)=1$, and

$$e_i=d_i(A)/d_{i-1}(A), \quad \text{for } i=1,2,\ldots, r; \qquad \ldots (2)$$

then $e_1, —, e_r$ are called the *invariant factors* of A. Also

$$d_i(A)=e_1e_2\ldots e_i, \quad \text{for } i=1,2,\ldots r. \qquad \ldots (3)$$

Hence the invariant factors of a matrix determine uniquely its determinant divisors, and conversely.

It now follows immediately from VI.3.2 and (2) that

VI.3.3. *If B is equivalent to a matrix A, of rank r, over a Euclidean domain, then A, B have the same invariant factors.*

Now it follows, from VI.3.1, that the determinant divisors of A, being the same as those of S, are given by (taking $d_i(A)=d_i$):

$$d_1=e_1,\ d_2=e_1e_2,\ \ldots,\ d_r=e_1e_2 \ldots e_r.$$

so that

$$e_1=d_1,\ e_2=d_2/d_1,\ \ldots,\ e_r=d_r/d_{r-1}.$$

Hence, it follows by (2), that:

VI.3.4. *The diagonal elements $e_1, e_2, \ldots, e_r$ in the Smith's normal form S of a matrix A are the invariant factors of A.*

Also, two matrices, equivalent over a Euclidean domain, have the same set of invariant factors. Conversely, two matrices having the same set of invariant factors, being equivalent to the same Smith's normal form S, are equivalent to one another. Thus one obtains :

VI.3.5. *The matrices* S, *defined in* VI.3.1 *and characterised by* VI.3.4, *constitute a complete set of cononical matrices for the set of all square matrices, of order* n, *under equivalence over a Euclidean domain.*

(ii) *Two square matrices of order* n *are equivalent over a Euclidean domain if, and only if, they have same set of invariant factors.*

EXAMPLE 2 : To find the determinant divisors, and the invariant factors of the following matrix :

$$B=\begin{pmatrix} x^2-x & x & 0 \\ x-2 & x & x+4 \\ x & 0 & 8x \end{pmatrix}.$$

The determinant divisors of B are given by :

$d_1=$ h.c.f. of $x^2-x,\ x,\ x-2,\ x,\ x+4,\ x,\ 3x=1$

$d_2=$ h.c.f. of $x(x^2-2x+2),\ x(x-1)(x+4),\ x(x+4),\ -x^2,\ 3x(x^2-x),\ 8x^2,$
$-x^2,\ 2x(x-5),\ 8x^2$

$=x$

$d_3=x^2(8x^2-5x+10)$.

The invariant factors of B are therefore given by

$$e_1=d_1=1,\quad e_2=d_2/d_1=x,\quad e_3=d_3/d_2=x(3x^2-5x+10).$$

Elementary divisors. Let $p_1,\ldots,p_m$ be the irreducible elements in the Euclidean domain W, each of which is a divisor of at least one of the invariant factors $e_1,\ldots,e_r$ of a matrix A.

Then $$e_k=p_1^{t_{k1}}\ldots\ldots p_m^{t_{km}},\ \ldots,\ k=1,2,\ldots,r,$$

where t_{ki} are non-negative integers. The prime-powers

$$p_1^{t_{11}},\ \ldots,\ p_m^{t_{1m}},\ \ldots,\ p_1^{t_{r1}},\ \ldots,\ p_m^{t_{rm}},$$

for which the indices are positive, are called the *elementary divisors of* A *over* W. It may be noted that whereas the invariant factors of a matrix, being given in terms of h.c.f. of its elements, remain invariant, its elementary divisors may not remain invariant if the ground system W be extended to a larger domain (since the elements p_i, irreducible in W, may be reducible in the extended domain).

For the elementary divisors, it is customary to write the powers of the same prime together in order of non-increasing indices. For instance, if the invariant factors of a matrix A, of order 7, when resolved into prime-power factors in a Euclidean domain W, be

$$e_1=1,\ e_2=1,\ e_3=p_1,\ e_4=p_1p_2,\ e_5=p_1^3p_2^2,$$
$$e_6=p_1^3p_2^2p_3p_4, \text{ and } e_7=p_1^5p_2^3p_3^2p_4p_5^4,$$

then the elementary divisors of A in W are written as follows :

$$p_1^5, p_1^3, p_1^3, p_1, p_1\ ;\ p_2^3, p_2^2, p_2^2, p_2\ ;\ p_3^2, p_3\ ;\ p_4, p_4\ ;\ p_5^4.$$

On the other hand, if the elementary divisors are given, as above, the invariant factors can readily be reconstituted. In fact, to form the last invariant factor e_7, one has to collect the highest powers of the distinct primes among the elementary divisors, and form their product, so that

$$e_7=p_1^5.p_2^3.p_3^2.p_4.p_5^4.$$

Let the prime powers selected for e_7, be struck off from the list of elementary divisors. Then the product of the highest powers of the distinct primes in the remaining elementary divisors gives e_6, so that

$$e_6=p_1^3.p_2^2.p_3.p_4.$$

The method is repeated till the elementary divisors are all exhausted. Thus, one obtains :

$$e_5=p_1^3.p_2^2,\quad e_4=p_1p_2,\quad e_3=p_1.$$

The two remaining invariant factors must be 1, that is,

$$e_2=1, \text{ and } e_1=1.$$

In forming the invariant factors from the elementary divisors, the order of the matrix and its rank must be known beforehand. For instance, in the converse part of the above example, if the order of the matrix A be 12 and its rank is 10, then with the same set of elementary divisors, of the invariant factors, now 12 in number, the last two must be 0, i. e., $e_{12}=0$, $e_{11}=0$, and the others would be given by

$$e_{10}=p_1^5p_2^3p_3^2p_4p_5^4,\quad e_9=p_1^3p_3^2p_3p_4,\quad e_8=p_1^3p_2^2,$$
$$e_7=p_1p_2,\quad e_6=p_1,\quad e_5=e_4=e_3=e_2=e_1=1.$$

Thus, the invariant factors of a matrix A in a Euclidean domain W are obtained from the elementary divisors, when the order and the rank of the matrix A are given. Hence, it follows immediately from VI.3.5(ii) that :

VI.3.6. *Two matrices of order n, over a Euclidean domain W, have the same set of invariant factors if, and only if, they have the same rank and the same set of elementary divisors.*

Now it follows immediately from VI.3.5(ii) that

VI.3.7. *Two matrices of order n are equivalent over a Euclidean domain W if, and only if, they have the same rank and the same set of elementary divisors in W.*

EXAMPLE **3** : To find the elementary divisors of

(i) the matrix A of example 1 in the ring of integers J, and

(ii) the matrix B of example 2 in R[x] and C[x], where R and C are respectively the fields of rational and complex numbers.

Soln : (i) The invariant factors of A, resolved into prime-power factors in J are

$$2=2,\quad 2=2,\quad 4=2^2,\ \text{and}\ 736=2^5.23.$$

Hence the elementary divisors of A in J are given by

$$2^5,\ 2^2,\ 2,\ 2\ ;\ 23,$$

(ii) The invariant factors of B, resolved into prime-power factors in R[x] are

$$e_1=1,\quad e_2=x,\quad e_3=3x^3-5x^2+10x = x(3x^2-5x+10).$$

Hence the elementary divisors of B in R[x] are given by

$$x,\ x\ ;\ 3x^2-5x+10.$$

But the invariant factors of B, resolved into prime-power factors in C[x] are

$e_1=1$, $e_2=x$, $e_3=3x^3-5x^2+10x=x(x-\alpha)(x-\beta)$, where α, $\beta=(5\pm i\sqrt{95})/6$.

Hence the elementary divisors of B in C[x] are given by

$$x,\ x\ ;\ x-\alpha\ ;\ x-\beta.$$

Exercises

1. Show that if a matrix A be equivalent to a diagonal matrix $D=(d_i)$, over a Euclidean domain W, then the prime-power factors of d_i are the elementary divisors of A over W.

2. Find the elementary divisors of the following matrices :

$$\begin{pmatrix} x+3 & 3 & 2x \\ -x & x^2-1 & x+1 \\ x-1 & x-1 & x-2 \end{pmatrix}, \quad \begin{pmatrix} 3 & 4 & -1 & 5 \\ 6 & 0 & 7 & -2 \\ 5 & -3 & 2 & 8 \\ -2 & 5 & 9 & 6 \end{pmatrix},$$

$$\begin{bmatrix} 2 & & & & \\ & 3 & & & \\ & & 4 & & \\ & & & 7 & \\ & & & & 8 \end{bmatrix}, \quad \begin{pmatrix} x^2+1 & x^3+x & 2x^3-x^2+x \\ x-1 & x^2+1 & x^2-2x+1 \\ x^2 & x^3 & 2x^3-x^2+1 \end{pmatrix},$$

$$\begin{bmatrix} x^2+1 & x^2+3x+3 & x^4+4x-2 & x^2+3 \\ x-2 & x-1 & x+2 & x-2 \\ 3x+1 & 4x+3 & 2x+2 & 3x+2 \\ x^2+2x & x^2+6x+4 & x^2+6x-1 & x^2+2x+3 \end{bmatrix}.$$

3. Find the invariant factors of a matrix A, of rank and order 6, whose elementary divisors are given by

(i) $3^5, 3^2, 3, 3;\ 5^2, 5^2, 5;\ 11^2, 11;\ 13^4$.

(ii) $(x-1)^3, (x-1)^2, (x-1)^2, (x-1);\ (x+2)^3, (x+2);\ (x-7)^4$.

CHAPTER VII

CONGRUENCE OF MATRICES

Let Λ_n be the set of all square matrices of order n over a field K. A matrix B is said to be *congruent* to a matrix A, over K, if there exists a non-singular matrix P such that $B=PAP^t$, where $A,B,P \varepsilon \Lambda_n$. The non-singular matrix P is a product of some elementary matrices, say $P=E_1 \ldots E_r$. Then

$$\begin{aligned} PAP^t &= (E_1 \ldots E_r)A(E_1 \ldots E_r)^t = (E_1 \ldots E_r)A(E_r{}^t \ldots E_1{}^t) \\ &= E_1 \ldots \{E_{r-1}(E_rAE_r{}^t)E_{r-1}{}^t\} \ldots E_1{}^t. \end{aligned}$$

Thus B is obtained from A by applying some congruent transformations. Each congruent transformation comprises an elementary row operation and the *corresponding* elementary column operation.

The relation of being congruent is an equivalence relation. In fact, (i) $A=IAI^t$, (ii) $B=PAP^t$ implies $A=QBQ^t$, where $Q=P^{-1}$, and (ii) $B=PAP^t$ and $C=RBR^t$ imply $C=(RP)A(RP)^t$; these prove the reflexive, symmetric and transitive properties, since $Q=P^{-1}$ and RP are non-singular matrices. The set Λ_n is consequently partitioned into disjoint classes, and one obtains the classification of the square matrices (of order n) with respect to congruence.

1. Congruence of symmetric matrices.

Let Γ_n be the subset consisting of all symmetric matrices of order n in Λ_n. For a symmetric matrix A, $B=PAP^t$ is also symmetric, as $B^t=(PAP^t)^t=PAP^t=B$. Hence in the partition of Λ_n into congruent classes, Γ_n is partitioned into congruent classes of symmetric matrices. Normal forms of classes of congruent symmetric matrices are obtained as follows :

VII.1.1. *A symmetric matrix A of order n, over a field K, whose characteristic $\neq 2$, is congruent to a diagonal matrix D*

the number of non-null elements in the diagonal of D being the same as the rank of A.

Proof: If A be the null matrix, then $D=A$ satisfies the required conditions. Next let A be a non-null matrix. If the diagonal elements in A be all null, then some $a_{rs} \neq 0$; and $B=I_{rs}(1)A[I_{rs}(1)]^t$ has a non-null element $b_{rr}=2a_{rs}$ in its diagonal. Next $C=I_{1r}B[I_{1r}]^t$ has its leading element $c_{11}=b_{rr} \neq 0$. Thus A is congruent to a matrix C, having a non-null element in $(1,1)$-position. Then the first column and the first row of C are swept out by applying some row and the corresponding column operations, and one obtains a congruent matrix,

$$G=PCP^t=\begin{pmatrix} c_{11} & O \\ O & G_1 \end{pmatrix},$$

where $\qquad P=I_{21}(-c_{12}\,c_{11}^{-1})\ldots I_{n1}(-c_{1n}.c_{11}^{-1}),$

and G_1 is a symmetric matrix of order $n-1$.

The same procedure is now followed in case of G_1. In fact the entire matrix G is considered, and row and column operations are applied to reduce G_1 without affecting the first row and the first column of G. Continuing in this way, one obtains, in a finite number of steps, a diagonal matrix D congruent to A.

Again since multiplication by non-singular matrices does not affect the rank of a matrix, rank(A)=rank (D)=number of non-null elements in the diagonal of D.

The above theorem may as well be stated in the following form.

VII. 1.2. *Corresponding to a given symmetric matrix A, over a field K, whose characteristic* $\neq 2$, *there exists a non-singular matrix P, such that* $PAP^t=D$ *is a diagonal matrix, the number of non-null diagonal elements of D being the same as the rank of A.*

The diagonal elements d_i are not uniquely determined by the given matrix, since the matrix $I_i(u)D[I_i(u)]^t$ is also a diagonal matrix congruent to A, having a different element, viz. ud_iu, in the ith diagonal position. Hence the diagonal matrices D are *not* canonical.

Congruence over the field of real numbers. Let the ground field be taken, in particular, to be **E**, the field of real numbers. Then one obtains a more simplified form of theorem VII.1.1.

VII.1.3. (i) *A real symmetric matrix A, of rank r, is congruent to a matrix*

$$G=\begin{pmatrix} I_m & & \\ & -I_{r-m} & \\ & & O \end{pmatrix}, \quad \ldots (1)$$

where I_m and I_{r-m} are identity sub-matrices of orders m and $r-m$ respectively.

(ii) *The integers r and m are invariants for the matrix A under congruence.*

Proof: (i) By VII.1.1, A is congruent to a diagonal matrix $D=(d_i)$, where r diagonal elements of D are non-null. By suitable interchanges of rows and interchanges of corresponding columns, the r non-null diagonal elements can be brought in the first r diagonal positions, and in such a way that the first m among these are positive and the following $r-m$ are negative. So, it may be assumed without any loss of generality, that $d_1, \ldots, d_m$ are positive and $d_{m+1}, \ldots, d_r$ are negative Let

$$Q=I_1(1/\sqrt{d_1})\ldots I_m(1/\sqrt{d_m})I_{m+1}(1/\sqrt{(-d_{m+1})})\ldots I_r(1/\sqrt{(-d_r)}),$$

then $QDQ^t=G$ is of the form (1).

Also G is congruent to D, and therefore to A.

(ii) Since the rank of a matrix remains invariant under multiplication by non singular matrices, $r=\text{rank}(A)=\text{rank}(G)$ is an invariant of A under congruence. The invariance of m, under congruence, will be proved in VII.2.2.

The integer m, as defined above, is called the *index* of the symmetric matrix A, and $m-(r-m)=2m-r$ is called the *signature* of A.

It now follows, from the invariance of r and m in (1), under congruence, that two congruent real symmetric matrices have the same rank and signature. Conversely two real symmetric matrices having the same rank and signature, being congruent to the same normal form (1), are congruent to one another. Thus one obtains :

VII.1.4.(i) *The matrices* (1), *for different values of* r *and* m, *where* $0 \leqslant m \leqslant r \leqslant n$, *constitute a complete set of canonical matrices for the set of all symmetric matrices of order* n, *over the field of real numbers, under congruence.*

(ii) *Two real symmetric matrices of order* n *are congruent if, and only if, they have the same rank and signature.*

EXAMPLE 1 : Let A be a real symmetric matrix given by

$$A=\begin{pmatrix} 2 & 4 & 1 \\ 4 & 9 & 4 \\ 1 & 4 & 3 \end{pmatrix}.$$

(i) To reduce A to its normal form under congruence ; and to find its rank and signature.

(ii) To find the non-singular matrix P which reduces A to its normal from (i. e., such that PAP^t gives the normal form of A).

Soln : The method of sweep-out is applied, as follows, to reduce A to its normal form,

$$I_{31}(-\tfrac{1}{2})I_{21}(-2)A=\begin{pmatrix} 2 & 4 & 1 \\ 0 & 1 & 2 \\ 0 & 2 & \frac{5}{2} \end{pmatrix}=A_1, \quad A_1 I_{12}(-2)I_{13}(-\tfrac{1}{2})=\begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & \frac{5}{2} \end{pmatrix}=B,$$

$$I_{32}(-2)B=\begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & -\frac{3}{2} \end{pmatrix}=B_1, \qquad B_1 I_{23}(-2)=\begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\frac{3}{2} \end{pmatrix}=C,$$

and $$I_3\left(\frac{\sqrt{2}}{\sqrt{3}}\right)I_1\left(\frac{1}{\sqrt{2}}\right).\,C.\,I_1\left(\frac{1}{\sqrt{2}}\right)I_3\left(\frac{\sqrt{2}}{\sqrt{3}}\right)=\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}=D.$$

D gives the normal form of A under congruence. Also

rank of A $=$ rank of $D=3$, and

signature of $A=$(the number of 1)$-$(the number of -1) in the diagonal of D

$=2-1=1$

Again $$D=I_3\left(\frac{\sqrt{2}}{\sqrt{3}}\right)I_1\left(\frac{1}{\sqrt{2}}\right)I_{32}(-2)I_{31}(-\tfrac{1}{2})I_{21}(-2).A.\,I_{12}(-2)I_{13}(-\tfrac{1}{2}) \\ I_{23}(-2)I_1\left(\frac{1}{\sqrt{2}}\right)I_3\left(\frac{\sqrt{2}}{\sqrt{3}}\right)$$

$=PAP^t$,

where $$P=I_3\left(\frac{\sqrt{2}}{\sqrt{3}}\right)I_1\left(\frac{I}{\sqrt{2}}\right)I_{32}(-2)I_{31}(-\tfrac{1}{2})\,I_{21}(-2)$$

$$=I_3\left(\frac{\sqrt{2}}{\sqrt{3}}\right)I_1\left(\frac{1}{\sqrt{2}}\right)I_{32}(-2)I_{31}(-\tfrac{1}{2})\begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$=I_3\left(\frac{\sqrt{2}}{\sqrt{3}}\right)I_1\left(\frac{1}{\sqrt{2}}\right)I_{32}(-2)\begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{pmatrix}$$

$$=I_3\left(\frac{\sqrt{2}}{\sqrt{3}}\right)I_1\left(\frac{1}{\sqrt{2}}\right)\begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ \frac{7}{2} & -2 & 1 \end{pmatrix}=\begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 \\ -2 & 1 & 0 \\ \frac{7\sqrt{2}}{2\sqrt{3}} & \frac{-2\sqrt{2}}{\sqrt{3}} & \frac{\sqrt{2}}{\sqrt{3}} \end{bmatrix}.$$

Exercises

1. Find a non-singular matrix P such that PAP^t is a diagonal matrix, none of whose diagonal elements is other than 1,—1, 0, where A is a real symmetric matrix given by

$$\begin{pmatrix} 2 & 1 & 8 \\ 1 & -3 & 9 \\ 8 & 9 & 2 \end{pmatrix}, \quad \begin{pmatrix} 0 & 2 & 3 \\ 2 & 4 & 0 \\ 3 & 0 & 1 \end{pmatrix},$$

$$\begin{bmatrix} 2 & -2 & -1 & \frac{11}{2} \\ -2 & -1 & 3 & 5 \\ -1 & 3 & -5 & 3 \\ \frac{11}{2} & 5 & 3 & 7 \end{bmatrix}, \quad \begin{bmatrix} 5 & -1 & 2 & 0 \\ -1 & 1 & 2 & 2 \\ 2 & 2 & 0 & \frac{1}{2} \\ 0 & 2 & \frac{1}{2} & 3 \end{bmatrix}.$$

2. Find the rank and signature of the matrices in exercise 1.

3. Find whether the following matrices are congruent :

$$\begin{pmatrix} -1 & 2 & 5 \\ 2 & 0 & -3 \\ 5 & -3 & 4 \end{pmatrix} \text{ and } \begin{pmatrix} 2 & -3 & 7 \\ -3 & 0 & 5 \\ 7 & 5 & 0 \end{pmatrix}.$$

2. Quadratic forms.

A homogeneous polynomial of degree 2 in n variables $x_1,\ldots,$ x_n, over a field K, is called a *quadratic form.* Let $\Sigma b_{ik}x_ix_k$. $i,k=1,2,\ldots,n$, be a quadratic form over K, and let the characteristic of K be different from 2. Putting $a_{ik}=(b_{ik}+b_{ki})/2$, the given quadratic form is transformed to $\Sigma a_{ik}x_ix_k$, where $a_{ik}=a_{ki}$. It can be expressed as a matrix product XAX^t, where X is a 1-row matrix $(x_1,\ldots,x_n)$ and A is the symmetric matrix (a_{ik}).

By a transformation $X=YP$, with a non-singular matrix P, the quadratic form XAX^t is transformed to $(YP)A(YP)^t=Y(PAP^t)Y^t$ in new variables $y_1, \ldots, y_n$; and the symmetric matrix A of the given quadratic form is transformed to a congruent symmetric matrix PAP^t.

Now, for a suitable non-singular matrix $P=(p_{ik})$, $PAP^t=D=(d_i)$ is a diagonal matrix (by VII.1.2). The non-null elements $d_1, \ldots, d_r$ in the diagonal of D are r in number, where r is the rank of A. Thus, one obtains :

VII.2.1. *A given quadratic form* $XAX^t=\Sigma a_{ik}x_ix_k$, $a_{ik}=a_{ki}$, $i,k=1,2,\ldots,n$, *over a field* K *of characteristic* $\neq 2$, *can be reduced to the diagonal form* $d_1y_1^2+\ldots+d_ry_r^2$, r *being the rank of* (a_{ik}), *by a suitable linear substitution* $X=YP$, *i.e.,* $x_i=\Sigma p_{ik}y_k$, $i,k=1,2,\ldots,n$, *where* $\det(p_{ik}) \neq 0$.

Particularly, over the field of real numbers, one obtains :

✓VII.2.2. (i) *Any given quadratic form* $XAX^t=\Sigma a_{ik}x_ix_k$, $a_{ik}=a_{ki}$, *with real coefficients, can be reduced to the diagonal form* $y_1^2+\ldots+y_m^2-y_{m+1}^2-\ldots-y_r^2$ *by a suitable linear substitution* $X=YP$, *i.e.,* $x_i=\Sigma p_{ik}y_k$, *where* $\det(p_{ik}) \neq 0$.

(ii) r *and* m *are invariants of the given quadratic form.*

Proof : Let XAX^t be a given real quadratic form, where $X=(x_1, \ldots, x_n)$, and $A=(a_{ik})$ is a symmetric matrix.

(i) By VII.1.3, A is congruent to a diagonal matrix G, given by (1), page 134 ; and $G=PAP^t$, where $P=(p_{ik})$ is a suitable non-singular matrix. The transformation $X=YP$, i.e., the substitution, $x_i=\Sigma p_{ik}y_k$, transforms the given quadratic form into its normal form YGY^t, i.e., into

$$y_1^2+\ldots+y_m^2-y_{m+1}^2-\ldots-y_r^2.$$

(ii) By VII.1.3(ii), r is an invariant of A under congruence. To prove the invariance of m under congruence, let A be also congruent to a second matrix.

$$H=\begin{pmatrix} I_k & & \\ & -I_{r-k} & \\ & & O \end{pmatrix},$$

where $k \neq m$. Without any loss of generality, it may be assumed that $k < m$.

Now $G=PAP^t$ and $H=QAQ^t$, where P and Q are suitable non-singular matrices, and the quadratic form XAX^t is transformed, by $X=YP$ and $X=ZQ$, into

$$YGY^t=y_1^2+\ldots+y_m^2-y_{m+1}^2-\ldots-y_r^2, \qquad \ldots (1)$$

and $$ZHZ^t=z_1^2+\ldots+z_k^2-z_{k+1}^2-\ldots-z_r^2. \qquad \ldots (2)$$

The quadratic forms (1) and (2) are equivalent to the form XAX^t under the substitutions

$$Y=XP^{-1}, \text{ i.e., } y_i=\Sigma u_{ik}\,x_k, \text{ where } P^{-1}=(u_{ik}),$$

and $$Z=XQ^{-1}, \text{ i.e., } z_i=\Sigma v_{ik}x_k, \text{ where } Q^{-1}=(v_{ik}).$$

Now the following linear homogeneous equations,

$$\left.\begin{array}{l} v_{11}x_1+\ldots\ldots\ldots+v_{1n}x_n=0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ v_{k1}x_1+\ldots\ldots\ldots+v_{kn}x_n=0 \\ u_{m+1\,1}x_1+\ldots+u_{m+1\,n}x_n=0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ u_{n1}x_1+\ldots\ldots\ldots+u_{nn}x_n=0 \end{array}\right\}, \qquad \ldots (3)$$

being $k+(n-m)=n-(m-k)<n$ in number, has a non-null solution $(h_1,\ldots,h_n)=X_0$, say. For $X=X_0$, the value of the given quadratic from is $X_0AX_0^t$. This value must be the same as $Y_0GY_0^t$ and $Z_0HZ_0^t$, where $Y_0=X_0P^{-1}$ and $Z_0=X_0Q^{-1}$. But by (3), the first k co-ordinates of Z_0 are null and the last $n-m$ co-ordinates of Y_0 are null, that is,

$$Y_0=(c_1,\ldots c_m, 0,\ldots,0) \text{ and } Z_0=(0,\ldots,0, g_{k+1},\ldots,g_n).$$

Then, by (1) and (2),

$$Y_0GY_0^t=c_1^2+\ldots+c_m^2 \text{ and } Z_0HZ_0^t=-g_{k+1}^2-\ldots-g_r^2.$$

Now, in order that these two values may be equal, one must have $c_1=\ldots=c_m=0$ and $g_{k+1}=\ldots=g_r=0$, so that Y_0, Z_0 are null vectors. But, then $X_0=Y_0P$ implies that X_0 is also the null vector, contrary to the fact thas $X_0=(h_1,\ldots,h_n)$ is a non-null vector satisfying the equations (3). Hence the initial assumption that $k \neq m$ is not valid. This proves the invariance of m.

The integer m, as defined above, is called the *index* of the given quadratic form. The number $m-(r-m)=2m-r$ is called the *signature* of the quadratic form.

The result (ii) in VII.2.2. is known as *Sylvester's law of inertia* for quadratic forms.

EXAMPLE 1 : To reduce the quadratic form

$$2x^2+9y^2+3z^2+8xy+2xz+8yz$$

to its normal form under congruence ; and to find its rank and signature.

Soln : The symmetric matrix of the quadratic form is the matrix A in example 1, page 135. The normal form of the matrix A, under congruence, has been found to be the matrix D in that example. The normal form of the given quadratic form is obtained from the matrix D as

$$X^2+Y^2-Z^2.$$

The rank and signature of the quadratic form are, accordingly, found to be 3 and 1 respectively.

Lagrange's method of reduction. The reduction of a quadratic form to its (diagonal) normal form can also be done by a method due to Lagrange, in the following way.

Let $\sum\limits_{i,j=1}^{n} a_{ji}x_ix_j$, $a_{ij}=a_{ji}$, be a given quadratic form, over a field K, whose characteristic $\neq 2$.

Case 1. Let one of the coefficients a_{ii} be non-null. Without any loss of generality, it may be assumed that $a_{11}\neq 0$. Then

$$\sum_{i,j=1}^{n} a_{ij}x_ix_j=a_{11}x_1^2+2x_1(a_{12}x_2+\dots+a_{1n}x_n)+\sum_{i,j=2}^{n} a_{ij}x_ix_j$$

$$=a_{11}^{-1}[a_{11}x_1+(a_{12}x_2+\dots+a_{1n}x_n)]^2+\sum_{i,j=2}^{n} b_{ij}x_ix_j,$$

where $\sum\limits_{i,j=2}^{n} b_{ij}x_ix_j=\sum\limits_{i,j=2}^{n} a_{ij}x_ix_j-a_{11}^{-1}(a_{12}x_2+\dots+a_{1n}x_n)^2$

is a quadratic form in the variables $x_2,\dots,x_n$. Repeating the process, one can express the given quadratic form as

$$\sum_{i,j=1}^{n} a_{ij}x_ix_j=d_1(g_{11}x_1+\dots+g_{1n}x_n)^2+d_2(g_{22}x_2+\dots+g_{2n}x_n)^2+$$
$$\dots+d_{n-1}(g_{n-1,n-1}x_{n-1}+g_{n-1,n}x_n)^2+d_n(g_{nn}x_n)^2$$
$$=d_1y_1^2+d_2y_2^2+\dots+d_ny_n^2,$$

by the substitution

$$\begin{aligned} g_{11}x_1+\ldots\ldots\ldots\ldots+g_{1n}x_n&=y_1,\\ g_{22}x_2+\ldots+g_{2n}x_n&=y_2,\\ \ldots\qquad\quad \ldots&\\ g_{nn}x_n&=y_n. \end{aligned}$$

Case 2. If all the coefficients a_{ii} be null, but some a_{ij}, say a_{12}, be non-null, then putting

$$x_1=y_1,\ x_2=y_1+y_2, \text{ and } x_i=y_i, \text{ for } i=3.\ldots..,n,$$

the given quadratic form is transformed to a quadratic form in the variables $y_1,\ldots.,y_n$, for which the coefficient of y_1^2 is $2a_{12}\neq 0$. One can then proceed as in case 1.

In fact, if in the process of successive reduction, the coefficients of the square terms of the remaining variables are all null, then a substitution as above is applied to obtain a square term with a non-null coefficient. And then the method, outlined in case 1, is followed.

If the ground field be, in particular, the field of real numbers, or any (formally) real field in general, then the (diagonal) normal form $d_1y_1^2+\ldots+d_ny_n^2$ (where some of the d's may be null) can be further reduced to the particular (diagonal) normal form, whose coefficients are 1, −1 or 0, by the substitution $z_i=\sqrt{|d|_i}\,y_i$ for $i=1,\ldots,n$.

EXAMPLE 2: To express the following quadratic forms as algebraic sums of squares, and to find their normal forms, ranks and signatures :

(i) $-2x_1^2-4x_1x_2+22x_2^2+6x_1x_3+12x_2x_3-x_3^2$,

(ii) $2x_1x_2-x_1x_3+6x_1x_4+8x_2x_3+x_2x_4-4x_3x_4$.

Soln : (i) Let Q represent the given quadratic form (i). Then

$$\begin{aligned} Q&=-\tfrac12[4x_1^2+2.2x_1(2x_2-3x_3)\}]+22x_2^2+12x_2x_3-x_3^2\\ &=-\tfrac12(2x_1+2x_2-3x_3)^2+\tfrac12(2x_2-3x_3)^2+22x_2^2+12x_2x_3-x_3^2\\ &=-\tfrac12(2x_1+2x_2-3x_3)^2+24x_2^2+6x_2x_3+\tfrac72x_3^2\\ &=-\tfrac12(2x_1+2x_2-3x_3)^2+\tfrac38(8x_2+x_3)^2+\tfrac{25}{8}x_3^2\\ &=-\left[\frac{1}{\sqrt2}\left(2x_1+2x_2-3x_3\right)\right]^2+\left[\frac{3}{2\sqrt6}\left(8x_2+x_3\right)\right]^2+\left(\frac{5\sqrt2}{4}x_3\right)^2. \end{aligned}$$

The substitution

$$y_1=\frac{3}{2\sqrt6}(8x_2+x_3),\quad y_2=\frac{5\sqrt2}{4}x_3,\quad y_3=\frac{1}{\sqrt2}(2x_1+2x_2-3x_3),$$

transforms the given quadratic form to its normal form

$$y_1^2+y_2^2-y_3^2.$$

The rank of the quadratic form is therefore 3, and its signature is $2-1=1$.

(ii) The given quadratic form,

$$Q=2x_1x_2-x_1x_3+6x_1x_4+8x_2x_3+x_2x_4-4x_3x_4,$$

does not contain any square term. To introduce a square term, a transformation,

$$x_1=y_1,\ x_2=y_1+y_2,\ x_3=y_3,\ x_4=y_4,$$

is applied ; and the quadratic form is transformed into

$$Q=2y_1^2+2y_1y_2+2y_1y_3+7y_1y_4+8y_2y_3+y_2y_4-4y_3y_4$$

$$=2(y_1+\tfrac{1}{2}y_2+\tfrac{1}{2}y_3+\tfrac{7}{4}y_4)^2-\tfrac{1}{2}y_2^2+2y_2y_3-\tfrac{5}{2}y_2y_4-\tfrac{1}{2}y_3^2-\tfrac{49}{8}y_4^2-\tfrac{15}{2}y_3y_4,$$

(completing a perfect square with all terms involving y_1)

$$=2z_1^2-\tfrac{1}{2}(y_2-2y_3+\tfrac{5}{2}y_4)^2+\tfrac{3}{2}y_3^2-\tfrac{25}{2}y_3y_4-3y_4^2,$$

(by putting $y_1+\frac{1}{2}y_2+\frac{1}{2}y_3+\frac{7}{4}y_4=z_1$, and completing a perfect square with all terms involving y_2)

$$=2z_1^2-\tfrac{1}{2}z_2^2+\tfrac{3}{2}(y_3-\tfrac{25}{6}y_4)^2-\tfrac{697}{24}y_4^2$$

(by putting $y_2-2y_3+\frac{5}{2}y_4=z_2$, and completing a perfect square with all terms involoving y_3)

$$=2z_1^2-\tfrac{1}{2}z_2^2+\tfrac{3}{2}z_3^2-\tfrac{697}{24}z_4^2 \text{ (by putting } y_3-\tfrac{25}{6}y_4=z_3, \text{ and } y_4=z_4)$$

$$=u_1^2+u_2^2-u_3^2-u_4^2$$

(by putting $\sqrt{2}z_1=u_1$, $\sqrt{\frac{3}{2}}\,z_3=u_2$, $\frac{1}{\sqrt{2}}z_2=u_3$ and $\sqrt{\frac{697}{24}}z_4=u_4$).

Thus the substitution,

$$\left.\begin{aligned}
u_1&=\sqrt{2}z_1=\sqrt{2}(y_1+\tfrac{1}{2}y_2+\tfrac{1}{2}y_3+\tfrac{7}{4}y_4)=\frac{\sqrt{2}}{4}(2x_1+2x_2+2x_3+7x_4)\\
u_2&=\frac{\sqrt{3}}{\sqrt{2}}z_3=\frac{\sqrt{6}}{12}(6y_3-25y_4)\qquad=\frac{\sqrt{6}}{12}(6x_3-25x_4)\\
u_3&=\frac{1}{\sqrt{2}}z_2=\frac{1}{2\sqrt{2}}(2y_2-4y_3+5y_4)\quad=\frac{\sqrt{2}}{4}(-2x_1+2x_2-4x_3+5x_4)\\
u_4&=\frac{\sqrt{697}}{\sqrt{24}}z_4=\frac{\sqrt{697}}{\sqrt{24}}y_4\qquad=\frac{\sqrt{4182}}{12}x_4
\end{aligned}\right\}.$$

transforms the given quadratic form into its normal form,

$$u_1^2+u_2^2-u_3^2-u_4^2.$$

Also $Q=\left[\frac{\sqrt{2}}{4}\left(2x_1+2x_2+2x_3+7x_4\right)\right]^2+\left[\frac{\sqrt{6}}{12}\left(6x_3-25x_4\right)\right]^2$

$$-\left[\frac{\sqrt{2}}{4}\left(-2x_1+2x_2-4x_3+5x^4\right)\right]^2-\left[\frac{\sqrt{4182}}{12}x_4\right]^2.$$

The rank of the quadratic form is 4, and its signature is $2-2=0$.

Exercises

1. Reduce the following quadratic forms to their normal forms, and find their ranks and signatures :

(i) $3x^2+7y^2+z^2-7xy-3yz$

(ii) $2x^2+3y^2-3xy-2xz+yz$

(iii) $x_1^2+25x_2^2+9x_3^2+10x_1x_2-6x_1x_3-30x_2x_3$

(iv) $3x_1^2-3x_2^2-8x_3^2+8x_1x_2-10x_1x_3+10x_2x_3$

(v) $2x_1^2-x_4^2-x_1x_3+2x_2x_3-x_1x_4$

(vi) $2x_1^2+3x_2^2+4x_3^2-8x_4^2+6x_1x_2+4x_2x_3+8x_1x_4$

(vii) $5x_1^2+3x_2^2-2x_3^2-9x_4^2-12x_1x_2+12x_1x_3-15x_1x_4$
$-2x_2x_3+6x_2x_4+2x_3x_4$

(viii) $3x_1^2+x_2^2-4x_3^2-11x_4^2-8x_1x_2+8x_1x_3-13x_1x_4$
$+2x_2x_3+2x_2x_4+2x_3x_4$

(ix) $2x_1^2-x_2^2-5x_3^2+7x_4^2-4x_1x_2-2x_1x_3+81x_1x_4$
$+6x_2x_3-10x_2x_4+6x_3x_4$

(x) $5x_1^2+x_2^2+3x_4^2-2x_1x_2+4x_1x_3+4x_2x_3$
$+4x_2x_4+x_8x_4$.

2. Express each of the quadratic forms in exercise 1 as an algebraic sum of squares.

3. Express $yz+zx+xy$ as the algebraic sum of three squares.

2.1. Definite, semi-definite and indefinite forms.

A real quadratic form $XAX^t=\Sigma\, a_{ij}x_ix_j$, $i, j=1,\ldots,n$, where $A=(a_{ii})$ is a symmetric matrix with rank r and index p, is said to be

(i) *positive definite*, if $r=p=n$,

(ii) *positive semi-definite*, if $r=p<n$,

(iii) *negative definite*, if $r=n$, $p=0$,

(iv) *negative semi-definite*, if $r<n$, $p=0$.

In all other cases, the form is said to be *indefinite*. Thus for an indefinite form, $0<p<r\leqslant n$.

It follows from VII.2.2 (i) that, by a suitable linear transformation $X=YP$, *i.e.*, $x_i=\sum\limits_{j=1}^{n}p_{ij}y_j$, $i=1,\ldots,n$, where $\det(p_{ij})\neq 0$, the above mentioned forms can be reduced to their respective normal (diagonal) forms as follows.

A positive definite form can be reduced to the normal form

$$y_1^2+\dots+y_n^2,$$

and can therefore represent only a positive real number (unless $x_1=\dots=x_n=0$).

A positive semi-definite form can be reduced to the normal form

$$y_1^2+\dots+y_r^2, \text{ where } r<n,$$

and can therefore represent either a positive real number, or the number 0 (even when all the x_i's are not 0).

A negative definite form can be reduced to the normal form

$$-y_1^2-\dots-y_n^2,$$

and can therefore represent only a negative real number (unless $x_1=\dots=x_n=0$.

A negative semi-definite form can be reduced to the normal form

$$-y_1^2-\dots-y_r^2, \text{ where } r<n,$$

and can therefore represent either a negative real number, or the number 0 (even when all the x_i's are not 0).

An indefinite form can be reduced the normal form

$$y_1^2+\dots+y_p^2-y_{p+1}^2-\dots-y_r^2, \text{ where } 0<p<r\leqslant n,$$

and can therefore represent a positive real number or a negative real number or the number 0 (even when all the x_i's are not 0).

Exercises

1. Verify that the form $5x^2+6y^2+14z^2-14yz-10zx$ is positive definite.

2. Find which of the quadratic forms in exercise 1, page 142, are definite (positive or negative), and which are indefinite ; find also the semi-definite forms.

(a) For the indefinite forms find two sets of values of the variables for which the form assumes positive and negative values.

(b) For every semi-definite form find a non-zero set of values of the variables which makes the form zero.

2.2. Application (Projective classification of quadrics).

Projective co-ordinates of points in the real n-dimensional projective space P^n are given by $(n+1)$-tuples of real numbers $(x_1,\dots,x_{n+1})$. The frame of reference, comprising the fundamental simplex and the unit point, consists of an ordered set of $n+1$ *independent* points $A_1,\dots,A_n$, U. The points $A_1,\dots,A_n$ are the successive vertices of the fundamental simplex, and U is the unit point. Referred to a different frame of reference, $B_1,\dots,B_n$, V, one obtains a different set of projective co-ordinates of the points. A change of the frame of reference induces a transformation of co-ordinates for the points, which is represented by an equation,

$$\rho X = YP,$$

where $X=(x_1,\dots,x_{n+1})$ and $Y=(y_1,\dots,y_{n+1})$ are the co-ordinates of the same point referred to the first and the second frames of reference respectively, and $P=(p_{ij})$ is a non-singular matrix of order $n+1$. The column-vectors of P are the co-ordinate vectors of the points B_1, $B_2,\dots$, B_n (forming the vertices of the second fundamental simplex) referred to the first frame of reference. Consequently, given any non-singular matrix P, one can choose the second frame of reference in such a way that the transformation of co-ordinates induced by the change from the first to the second frame of reference will be represented by the equation $\rho X = YP$.

A locus represented by a homogeneous equation of the second degree in P^n is called a hyper-quadric. Thus a hyper-quadric is represented by an equation

$$XAX^t = 0, \text{ i.e., } \Sigma\, a_{ij}x_ix_j = 0,\ i, j = 1,\dots, n+1,$$

where $A=(a_{ij})$ may be taken to be a symmetric matrix. The quadratic form XAX^t can be reduced to the diagonal form,

$$y_1^2 + \dots + y_p^2 - y_{p+1}^2 - \dots - y_r^2,$$

by a substitution $X=YP$, where P is a suitable non-singular matrix (vide VII.2.2 (i)). Corresponding to the non-singular matrix P, one can choose a frame of reference, such that a change from the initial frame of reference to this new one induces a transformation of co-ordinates of points, represented by the equation $\rho X=YP$. Hence, referred to the second frame of reference, the equation of the hyperquadric is reduced to the normal form

$$y_1^2+\dots+y_p^2-y_{p+1}^2-\dots-y_r^2=0. \qquad \dots (1)$$

Since r and p are invariaants, by the Sylvester's law of inertia, one obtains projectively distinct hyperquadrics corresponding to different values of r and p, or, equivalently, corresponding to different values of r and $s=p-(r-p)$, i.e., rank and signature. Without any loss of generality, it may be assumed that $p \geqslant r-p$ in (1), so that $s \geqslant 0$. Thus all the projectively distinct hyperquadrics are given in their normal forms by the equation (1), where r and $s=2p-r$ assume all values subject to the condition $0 \leqslant s \leqslant r \leqslant n+1$.

Projective classification of quadrics in 3-dimensional real projective space P^3. For $n=3$, i.e., in P^3, a hyperquadric is known as a quadric. Projectively distinct quadrics are given, in their normal forms, by the equation (1) for all values of r and s, subject to the condition $0 \leqslant s \leqslant r \leqslant 4$, as in the following table :

r	s	Normal form	Description
1	1	$y_1^2=0$	A pair of coincident planes.
2	2	$y_1^2+y_2^2=0$	A line quadric, $\left.\begin{array}{l} y_1=0 \\ y_2=0 \end{array}\right\}$.
2	0	$y_1^2-y_2^2=0$	A pair of distinct planes, $y_1+y_2=0$ and $y_1-y_2=0$.
3	3	$y_1^2+y_2^2+y_3^2=0$	A point quadric representing the point (0, 0, 0, 1).
3	1	$y_1^2+y_2^2-y_3^2=0$	A quadric cone.
4	4	$y_1^2+y_2^2+y_3^2+y_4^2=0$	A non-degenerate quadric without any real trace.
4	2	$y_1^2+y_2^2+y_3^2-y_4^2=0$	A non-degenerate non-ruled real quadric.
4	0	$y_1^2+y_2^2-y_3^2-y_4^2=0$	A non-degenerate ruled real quadric.

Projective classification of conics in real projective plane P^2.

For $n=2$, i.e., in P^2, a hyper-quadric is called a conic. Projectively distinct conics are given, in their normal forms, by the equation (1) for all values of r and s, subject to the condition $0 \leqslant s \leqslant r \leqslant 3$, as in the following table :

r	s	Normal form	Description
1	1	$y_1^2=0$	A pair of coincident lines.
2	2	$y_1^2+y_2^2=0$	A point conic representing the point (0, 0, 1).
2	0	$y_1^2-y_2^2=0$	A pair of distinct lines, $y_1+y_2=0$ and $y_1-y_2=0$.
3	3	$y_1^2+y_2^2+y_3^2=0$	A non-degenerate conic without any real trace.
3	1	$y_1^2+y_2^2-y_3^2=0$	A non-degenerate conic with real trace.

EXAMPLE 3 : Find the nature of the conic in P^2 represented by the equation $-2x_1^2+22x_2^2-x_3^2-4x_1x_2+6x_1x_3+12x_2x_3=0$.

Soln. It has been found in example 2 (i) that the rank and signature of the quadratic form on the left hand side of the given equation are 3 and 1 respectively. Hence the equation of conic in its normal form is given by $y_1^2+y_2^2-y_3^2=0$; it represents a non-degenerate conic with real trace.

EXAMPLE 4 ; Find the nature of the following quadric in P^3 :

$$2x_1x_2-x_1x_3+6x_1x_4+3x_2x_3+x_2x_4-4x_3x_4=0.$$

Soln. It has been found in example 2(ii) that the rank and signature of the quadratic form on the left hand side of the given equation are 4 and 0. Hence the equation of the quadric in its normal form is $y_1^2+y_2^2-y_3^2-y_4^2=0$; and it represents a non degenerate ruled quadric.

Exercises

1. Show that $\dfrac{x_1^2}{a^2}-\dfrac{x_2^2}{b^2}-2cx_3x_4=0$, $abc \neq 0$, can be transformed by a collineation into $x_1^2+x_2^2-x_3^2-x_4^2=0$.

2. Find the nature of the following quadrics in P^3 :

(i) $x_1^2-2x_1x_2+2x_2x_3-2x_3x_4+x_4^2=0$.

(ii) $2x_1^2 - x_4^2 - x_1x_2 + 2x_1x_3 - x_1x_4 = 0.$

(iii) $4x_1^2 - x_2^2 - x_3^2 - x_4^2 + 2x_2x_3 - 2x_2x_4 + 2x_3x_4 = 0.$

(iv) $2x_1^2 + 3x_2^2 + 4x_3^2 - 8x_4^2 + 6x_1x_2 + 4x_2x_3 + 8x_1x_4 = 0.$

(v) $5x_1^2 + x_2^2 + 3x_4^2 - 2x_1x_2 + 4x_1x_4 + 2x_2x_4 + 2x_3x_4 = 0.$

(vi) $x_1^2 + x_1^2 + x_3^2 + x_4^2 + 2x_2x_3 + 2x_3x_1 + 2x_1x_2 + 2x_1x_4$
$+ 2x_2x_4 - 2x_3x_4 = 0.$

(vii) $x_1^2 + 2x_2^2 + x_3^2 + x_4^2 + 3x_1x_2 - 2x_1x_3 + 2x_1x_4 - 3x_2x_3$
$+ 3x_2x_4 - 2x_3x_4 = 0.$

(viii) $9x_1^2 - 6x_2^2 - 8x_3^2 + 6x_1x_2 - 14x_1x_3 + 18x_1x_4 + 8x_2x_3$
$+ 12x_2x_4 - 4x_3x_4 = 0.$

2.3. Congruence of symmetric bilinear forms.

Let XAY^t be a bilinear form, where $A = (a_{ij})$ is an $m \times n$ matrix. When $m = n$, and A is symmetric, the bilinear form is called a *symmetric bilinear form.*

Let XAY^t be a given symmetric bilinear form. X and Y are said to be *transformed cogradiently*, if they are subjected to the same transformation, $X = X'P$ and $Y = Y'P$. The bilinear form is then transformed to $(X'P)A(Y'P)^t = X'(PAP^t)Y'^t$, which is also a symmetric bilinear form in new sets of variables.

It then follows from VII.1.2 and VII.1.3 that

VII.2.3. *Any symmetric bilinear form of rank r, over a field whose characteristic $\neq 2$, can be reduced to the normal (diagonal) form $d_1x_1y_1 + \ldots + d_rx_ry_r$, by a suitable cogradient transformation.*

VII.2.4. *Any real symmetric bilinear form can be reduced, by a suitable cogradient transformation, to the normal form,*

$$x_1y_1 + \ldots + x_py_p - x_{p+1}y_{p+1} - \ldots - x_ry_r,$$

where r and p are the rank and the index of the (symmetric) matrix of the given bilinear form.

Exercise

1. Reduce the symmetric bilinear form, $2x_1y_1 - x_1y_2 - x_2y_1 + x_2y_2 + 5x_1y_3 + 5x_3y_1 - 7x_3y_3$, to its normal form, under congruence.

3. Congruence of skew-symmetric matrices.

Any matrix, congruent to a skew-symmetric matrix, is skew-symmetric. As in the case of symmetric matrices, the set of skew-symmetric matrices of order n, over a field K, is partitioned into disjoint classes of congruent skew-symmetric matrices. The normal form is given by :

VII.3.1. *A skew-symmetric matrix A, of order n, over a field K whose characteristic $\neq 2$, is congruent to a matrix*

$$S=\begin{bmatrix} E_1 & & & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & E_s & \\ & & & & O \end{bmatrix}, \qquad \dots (1)$$

where $E_i=\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, *for* $i=1,2,\dots,s$.

Proof : If A be the null matrix, then $S=A$ satisfies the required conditions. Next, let $A=(a_{ik})$ be a non-null matrix, and let $a_{ik}=-a_{ki}=c\neq 0$. Then

$$B=I_1(c^{-1})I_{2k}I_{1i}AI_{1i}I_{2k}I_1'(c^{-1})$$

is congruent to A, and of the form

$$B=\begin{pmatrix} E_1 & B_2 \\ B_3 & B_4 \end{pmatrix}, \text{ where } E_1=\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

The element -1 in E_1 may now be utilised for sweeping out the first column of B by row operations, and then the element 1 in E_1 to sweep out the first row of B by the corresponding column operations. Next, the second column and the second row of B are swept out by some row and the corresponding column operations with the help of the elements 1 and -1 in E_1. Thus A is congruent to a matrix

$$C=\begin{pmatrix} E_1 & O \\ O & A_1 \end{pmatrix}.$$

Then C, and hence also A_1, are skew-symmetric. If A_1 be the null matrix, then $S=C$. Otherwise the process is repeated, and one obtains, in a finite number of steps, a matrix S with the required properties.

Again since the first $2s$ rows of S are linearly independent, and the others are null, rank of S is $2s$. Hence the rank of A, being the same as that of S, is also $2s$.

Corollary : The rank of a skew-symmetric matrix is an even number.

Now, since the normal form S, in VII.3.1. is determined, completely by the rank of the given matrix A, it follows that :

VII.3.2. (i) *The matrices* (1), *for* $s=0,1,2,\ldots, r$, *where* r *is the largest integer* $\leqslant n/2$, *constitute a complete set of canonical matrices for the set of skew-symmetric matrices of order* n, *over a field* K *whose characteristic* $\neq 2$.

(i) *Two skew-symmetric matrices, of order* n, *over a field* K *whose characteristic* $\neq 2$, *are congruent if, and only if, they have the same rank.*

Exercises

1. Reduce the following matrices to their normal forms, under congruence :

$$\begin{bmatrix} 0 & 2 & -1 & 0 & 3 \\ -2 & 0 & 4 & 1 & -2 \\ 1 & -4 & 0 & -5 & 6 \\ 0 & -1 & 5 & 0 & 1 \\ -3 & 2 & -6 & -1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 & 1 & 1 & 4 \\ 1 & 0 & -2 & 0 & 3 \\ -1 & 2 & 0 & 5 & -2 \\ -1 & 0 & -5 & 0 & -5 \\ -4 & -3 & 2 & 5 & 0 \end{bmatrix}.$$

Are the two matrices congruent ?

2. Show that any skew-symmetric matrix of rank $2s$ is congruent to a matrix

$$\begin{pmatrix} O & I_s & O \\ -I_s & O & O \\ O & O & O \end{pmatrix},$$

where I_s is the identity sub-matrix of order s.

4. Hermitian congruence

Let the ground field be taken to be **C**, the field of complex numbers. Hermitian congruence of Hermitian and skew-Hermitian matrices is a generalisation of the concept of congruence as defined for symmetric and skew-symmetric matrices over the field of real numbers. In fact, a matrix B of order n, over **C**, is said to be *Hermitely congruent* or *conjunctive* to a matrix A of order n, over **C**, if there exists a non-singular matrix P such that $B=PAP^{*t}$, where P^* denotes the matrix obtained by replacing every element in P by its conjugate. Then B is obtained from A by elementary row and column operations, in such a way that for every elementary row operation, a *conjunctively corresponding* column operation is also applied. Pairs of *conjunctively corresponding* elementary matrices are given by $I_k(c)$, $I_k(c^*)$ and $I_{ik}(c)$, $I_{ki}(c^*)$ where c^* is the conjugate of c.

It can be readily proved, as in the previous cases, that the relation of being Hermitely congruent is an equivalence relation ; and the set of all square matrices of order n over **C**, is accordingly partitioned into disjoint classes of Hermitely congruent matrices.

Hermitian matrices. A matrix, which is Hermitely congruent to a Hermitian matrix, is necessarily Hermitian. The set of all Hermitian matrices is partitioned into disjoint classes of Hermitely congruent Hermitian matrices.

Now, proceeding exactly in the same way as in the case of real symmetric matrices, one obtains, for the Hermitian matrices, the very same results as in VII.1.2 (i) & (ii) and VII.1.3 (i) & (ii). In fact,

VII.4.1. (i) *A Hermitian matrix A, of order n, is Hermitely congruent to a matrix*

$$G=\begin{pmatrix} I_p & & \\ & -I_{r-p} & \\ & & O \end{pmatrix}, \qquad \dots (1)$$

where I_p and I_{r-p} are identity sub-matrices of orders p and $r-p$ respectively.

(ii) *The integers r and p are invariants for the matrix A, under Hermitian congruence.*

VII.4.2. (i) *The matrices* (1), *for different values of r and p, where $0 \leqslant p \leqslant r \leqslant n$, constitute a complete set of canonical matrices for the set of all Hermitian matrices of order n, under Hermitian congruence.*

(ii) *Two Hermitian matrices of order n are Hermitely congruent if, and only if, they have the same rank and signature.*

Hermitian forms. A *Hermitian form*, as in the case of a quadratic form over the real field, is given by $\Sigma a_{ik} x_i x_k^*$, where $a_{ik} = a_{ki}^*$, and therefore also by a matrix product XAX^{*t}, where $A = (a_{ik})$ is a Hermitian matrix. One obtains accordingly the same results as in VII.2.4 (i) & (ii), viz.

VII.4.3. (i) *Any given Hermitian form $XAX^{*t} = \Sigma a_{ik} x_i x_k^*$, $a_{ik} = a_{ki}^*$, can be reduced to the normal diagonal form*

$$y_1^2 + \ldots + y_p^2 - y_{p+1}^2 - \ldots - y_r^2,$$

by a linear substitution $X = YP$, ie, $x_i = \Sigma p_{ik} y_k$, where $\det(p_{ik}) \neq 0$.

(ii) *The rank r and the signature $2p - r$ are invariants of the given Hermitian form*

Skew-Hermitian matrices. A matrix, Hermitely congruent to a skew-Hermitian matrix, is also skew-Hermitian. The set of skew-Hermitian matrices of order n is partitioned into disjoint classes of Hermitely congruent skew-Hermitian matrices. One does *not* however obtain results similar to VII.2.6 & VII.2.7(i) & (ii), which hold in the case of real skew-symmetric matrices.

To obtain the normal forms, the results for Hermitian matrices can however be utilised, since for a skew-Hermitian matrix S, the matrices iS and $-iS$ are both Hermitian.

Let $-iS = H$, then there exists a non-singular matrix P such that $PHP^{*t} = G$ is of the form (1). Then

$$iG = i.PHP^{*t} = P(iH)P^{*t} = PSP^{*t}.$$

Thus, one obtains the following :

VII.4.4. (i) *A skew-Hermitian matrix S, of order n, is Hermitely congruent to a matrix*

$$Q=\begin{pmatrix} iI_p & & \\ & -iI_{r-p} & \\ & & O \end{pmatrix}, \qquad \dots \quad (2)$$

where I_p *and* I_{r-p} *are identity sub-matrices of orders p and r−p respectively.*

(ii) *The integers r and p are invariants for the matrix S, under Hermitian congruence.*

It now follows immediately that

VII.4.5. (i) *The matrices* (2), *for different values of r and p, where* $0 \leqslant p \leqslant r \leqslant n$, *constitute a complete set of canonical matrices for the set of all skew-Hermitian matrices of order n, under Hermitian congruence.*

(ii) *Two skew-Hermitian matrices A, B, of order n, are Hermitely congruent if, and only if, they have the same rank, and iA and iB have the same signature.*

Exercises

1. Reduce the following Hermitian and skew-Hermitian matrices to their respective normal forms under Hermitian congruence :

$$\begin{pmatrix} -5 & 3-i & 2i \\ 3+i & 0 & 6 \\ -2i & 6 & 7 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1+i & -3 \\ -1+i & 0 & 4i \\ 3 & 4i & 0 \end{pmatrix}.$$

2. Prove that a Hermitian form always represents a real number.

3. Determine whether the Hermitian forms

$$3x_1x_1^* + 5ix_1x_2^* - 5ix_1^*x_2 - 2x_2x_2^*,$$

and $$(1+2i)x_1x_2^* + (1-2i)x_1^*x_2$$

are Hermitely congruent.

CHAPTER VIII

EIGEN VALUES AND EIGEN VECTORS

1. Matric polynomials.

Let Λ_n be the total matric algebra over a field K. Polynomials belonging to $\Lambda_n[x]$, i.e., having matrices as coefficients, are called *matric polynomials*. Any such matric polynomial,

$$F(x)=F_m x^m+\ldots+F_1 x+F_0, \qquad \ldots \quad (1)$$

can be expressed as a matrix over $K[x]$, i.e., a matrix with polynomial elements. Conversely, any matrix over $K[x]$ can be expressed as a matric polynomial over Λ_n.

The matric polynomial (1) is said to be *proper* if F_m is non-singular.

EXAMPLE 1 : $$\begin{pmatrix} 2 & -3 \\ 0 & 1 \end{pmatrix} x^2+\begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} x+\begin{pmatrix} 0 & 5 \\ 4 & 1 \end{pmatrix}$$
$$=\begin{pmatrix} 2x^2-x & -3x^2+5 \\ 4 & x^2+1 \end{pmatrix}.$$

Division algorithm. Division algorithm may also be defined for the matric polynomials. In fact,

VIII.1.1. *If $A(x)$ and $B(x)$ be two matric polynomials, of which $B(x)$ is proper and of degree k, then there exist unique matric polynomials $Q_l(x)$, $R_l(x)$ and $Q_r(x)$, $R_r(x)$, such that*

$$A(x)=B(x)Q_l(x)+R_l(x) \text{ and } A(x)=Q_r(x)B(x)+R_r(x),$$

where, either $R_l(x)$ and $R_r(x)$ are null polynomials, or their degrees are less than k.

Proof: Let $A(x)$ be of degree m. If $m<k$, then $Q_l(x)=Q_r(x)=0$ and $R_l(x)=R_r(x)=A(x)$ satisfy the required conditions. If, on the other hand, $m\geqslant k$, the matric polynomials $Q_l(x)$, $R_l(x)$ and $Q_r(x)$, $R_r(x)$ are obtained by the usual method of successive division algorithms, as for ordinary polynomials.

To prove the uniqueness of the quotients $Q_l(x)$, $Q_r(x)$ and the remainders $R_l(x)$, $R_r(x)$, let

$$A(x)=B(x)Q_l(x)+R_l(x)=B(x)S(x)+T(x).$$

Then $$B(x)[Q_l(x)-S(x)]=T(x)-R_l(x). \qquad \dots (2)$$

Now, either $Q_l(x)-S(x)=0$ and then $T(x)=R_l(x)$, or else the degree of the left hand side in (2) is $\geqslant k$, while that of the righth and side is $<k$ —which is impossible. Hence $Q_l(x)=S(x)$ and $R_l(x)=T$.

Uniqueness of $Q_r(x)$ and $R_r(x)$ is also similarly proved.

If now $B(x)$ be proper and of degree 1, one obtains :

VIII.1.2. *If* $A(x)=A_mx^m+\dots+A_1x+A_0$ *and* $B(x)=Ix-C$ *be matric polynomials over* Λ_n, *then the left and the right remainders, of division of* $A(x)$ *by* $B(x)$, *are given by*

$$\left.\begin{aligned} R_l&=C^mA_m+\dots+CA_1+A_0, \\ \textit{and} \quad R_r&=A_mC^m+\dots+A_1C+A_0 \end{aligned}\right\}. \qquad \dots (3)$$

Proof : $\Sigma A_ix^i-\Sigma C^iA_i=\Sigma(x^iI-C^i)A_i$

$$=\Sigma[(xI-C)(x^{i-1}I+x^{i-2}C+\dots+xC^{i-2}+C^{i-1})]A_i$$

$$=(xI-C)\,Q_l(x).$$

Thus $A(x)=(xI-C).Q_l(x)+R_l$, where R_l is given by (3).

The result can similarly be proved for R_r.

Matrix as a root of a polynomial. A matrix $M \in \Lambda_n$ is said to be a *root* of a polynomial $f(x)=a_nx^n+\dots+a_1x+a_0 \in K[x]$, if

$$f(M)=a_nM^n+\dots+a_1M+a_0I = \text{the null matrix.}$$

It will be shown that

(i) for any given square matrix M, over a field K, there exists a polynomial $f(x) \in K[x]$ having M as a root (vide VIII.2.1); and that

(ii) given any polynomial $f(x) \in K(x)$, there exists a square matrix M, over K, such that M is a root of $f(x)$ (vide VIII.2.2).

Let $f(xI)$ donote the matric polyomial $a_nx^nI+\dots+a_1xI+a_0I$.

then by VIII.1.2, the left and the right remainders of division of $f(xI)$ by $Ix-M$ are given by

$$M^n.\, a_nI+\ldots.+M.\, a_1I+a_0I=a_nM^n+\ldots+a_1M+a_0I=f(M),$$

and $$a_nI.M^n+\ldots+a_1I.M+a_0I=a_nM^n+\ldots+a_1M+a_0I=f(M).$$

Hence, it follows, by VIII.1.1, that :

VIII.1.3. *If $f(x)$ be a polynomial over a field K, then a matrix M, over K, is a root of $f(x)$ if, and only if, the matric polynomial $f(xI)$ is divisible by the matric polynomial $xI-M$ from either side i.e.,*

$$f(xI)=(xI-M).Q_l(x), \quad or \quad f(xI)=Q_r(x).(xI-M),$$

where $Q_l(x)$ and $Q_r(x)$ are suitable matric polynomials.

EXAMPLE 2: The matrix A, given in example 1, page 160, is a root of the polynomial x^2-2x+1. In fact,

$$A^2-2A+I=\begin{pmatrix}1&2&0\\0&1&0\\0&2&1\end{pmatrix}-\begin{pmatrix}2&2&0\\0&2&0\\0&2&2\end{pmatrix}+\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}=\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}.$$

Exercises

1. Express the following matrices as matric polynomials :

$$\begin{pmatrix}x^3-2x^2+5x+1 & x+3\\ x^2+5 & x^3-7x\\ 3x^3+2x-7 & x^2+2x-3\end{pmatrix}, \quad \begin{pmatrix}x^3-1 & 7x^2+3x-1\\ 2x^2+7x & x^3+x+1\\ x & 5\end{pmatrix}.$$

2. Find the left quotient and the left remainder, and the right quotient and the right remainder obtained by the division of $A(x)$ by $B(x)$, where

$$A(x)=\begin{pmatrix}5&-2\\3&1\end{pmatrix}x^3+\begin{pmatrix}1&-1\\2&3\end{pmatrix}x^2-\begin{pmatrix}2&0\\5&-1\end{pmatrix},$$

and $$B(x)=\begin{pmatrix}2&1\\3&2\end{pmatrix}x^2-\begin{pmatrix}1&0\\-2&1\end{pmatrix}x+\begin{pmatrix}2&3\\1&5\end{pmatrix}.$$

Q7 2. Characteristic polynomial.

Let $A=(a_{ij})$ be a matrix of order n over a field K. Then $A-xI$ is called the *characteristic matrix* for A. Det $(A-xI)$ is called the *characteristic polynomial* of A; and, correspondingly, det $(A-xI)=0$ is called the *characteristic equation* of the matrix A. The characteristic polynomial of A will be donoted by $\psi_A(x)$, so that

$$\psi_A(x)=\det(A-xI)=\begin{vmatrix} a_{11}-x & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22}-x & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn}-x \end{vmatrix}.$$

Thus, for the characteristic polynomial $\psi_A(x)$, one finds that:

(i) the leading coefficient is $(-1)^n$;

(ii) the coefficient of x^{n-1} is
$(-1)^{n-1}.(a_{11}+a_{22}+\dots+a_{nn})=(-1)^{n-1}.$ *trace of* A;

(iii) the coefficient of x^r is
$(-1)^r.$(sum of the principal minors of order $n-r$ of A),
[for $r=0, 1,\dots,n-1$]; and, in particular,

(iv) the constant term (i.e, the term indepedent of x) is det (A).

An important property of the characteristic polynomial is given by the following:

Q7 **VIII.2.1. Cayley-Hamilton theorem:** *Every square matrix is a root of its characteristic polynomial.*

Proof: Let $A=(a_{ij})$ be a square matrix of order n over a field K, and let $\psi_A(x)$ be its characteristic polynomial. Then $\psi_A(x)=\det(A-xI)$.

Let $P(x)$ denote the adjoint of the characteristic matrix $A-xI$. Then, by IV.2.3,

[∵ A·adj(A) = adj(A)·A = det|

$$(A-xI).\,P(x)=\det(A-xI).\,I=\psi_A{}'(x).\,I=\psi_A(xI).$$

Hence, A is a root of $\psi_A(x)$, by VIII.1.3.

The above theorem proves that every square matrix is a root of a polynomial (over the same ground field). Conversely, given any polynomial $f(x)$ of degree n, belonging to $K[x]$, where K is a field, there exists a square matrix of order n, having $(-1)^n f(x)$ as its characteristic polynomial. In fact, it can be easily verified by direct computation that :

VIII.2.2. *For any given polynomial,* $f(x)=x^n+b_{n-1}x^{n-1}+\ldots+b_1x+b_0$, *over a field* K, $(-1)^n.f(x)$ *is the characteristic polynomial of the following square matrix of order* n,

$$C(f)=\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ -b_0 & -b_1 & -b_2 & \cdots & -b_{n-2} & -b_{n-1} \end{bmatrix}. \quad \ldots (1)$$

The matrix $C(f)$, given by (1), is known as the *companion matrix* of the (monic) polynomial $f(x)$.

It is to be noted that different matrices may have the same characteristic polynomial (vide examples 1(ii), (iii), page 171).

EXAMPLE 1 : Express A^{-1} as a polynomial in A, where A is the matrix given in example 1(i), page 171.

$A=\begin{pmatrix} 3 & 2 & 2 \\ 1 & 4 & 1 \\ -2 & -4 & - \end{pmatrix}$

Soln. The characteristic polynomial of A is $-x^3+6x^2-11x+6$. Hence by Cauchy-Hamilton theorem,

$$-A^3+6A^2-11A+6I=0, \quad \text{i.e.,} \quad 6I=A^3-6A^2+11A,$$

or,

$$A^{-1}=\tfrac{1}{6}A^2-A+\tfrac{11}{6}I.$$

Exercises

1. Find the characteristic polynomials of the following matrices :

$$\begin{pmatrix} 2 & -1 & 3 \\ 0 & 5 & 4 \\ -3 & 2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & -7 \\ 3 & 0 & -5 \end{pmatrix}, \begin{pmatrix} 3 & 1 & 5 \\ 2 & -7 & 4 \\ -1 & 6 & 8 \end{pmatrix}.$$

2. For each of the following equations find a matrix having the given equation as its characteristic equation :

(a) $x^3-3x^2+4x+1=0$, (b) $2x^3-5x+6=0$,

(c) $x^4+2x^3-x^2+3=0$, (d) $3x^4-2x^3+6x^2+x=0$.

3. Minimal polynomial.

It has been proved in VIII.2.1 that for any square matrix A, over a field K, there exists a non-null polynomial in $K[x]$ having A as a root. A polynomial $m(x)$ of the least degree, having A as a root, is called a *minimal polynomial of* A *over* K.

VIII.3.1. *If A be a square matrix, over a field K, then A is a root of a polynomial $f(x) \in K[x]$ if, and only if, $f(x)$ is divisible by $m(x)$ in $K[x]$, where $m(x)$ is a minimal polynomial of A over K.*

Proof: As K is a field, $K[x]$ is a Euclidean domain. Hence, by division algorithm, one can express $f(x)$ as

$$f(x)=m(x).\,q(x)+r(x),$$

where the degree of $r(x) <$ the degree of $m(x)$.

Substituting A for x, one obtains,

$$f(A)=m(A).q(A)+r(A),$$

so that $\quad r(A)=f(A), \quad$ since $m(A)=O$.

Thus $f(A)=O$ if, and only if, $r(A)=O$.

On the other hand, if $f(x)$ be divisible by $m(x)$ in $K[x]$, then $r(x)$ is the null polynomial Then $r(A)=O$, and therefore $f(A)=O$. Hence A must be a root of $f(x)$.

It should however be noted that *a minimal polynomial need not be irreducible in* $K[x]$ (vide example 1).

It now follows immediately from VIII.2.1 and VIII.3.1 that :

VIII.3.2. *For any square matrix, over a field, the characteristic polynomal is divisible by the minimal polynomial.*

Again, since two minimal polynomials of the same matrix are mutually divisible, it follows that they are associated in $K[x]$, where K is the ground field. Thus *a minimal polynomial of a matrix, over a field K, is determined uniquely, upto a unit factor in $K[x]$*; hence one uses the expression "*the*" minimal polynomial of a matrix.

VIII.3.3. *The minimal polynomial of a square matrix A, over a field K, is the last invariant factor of the characteristic matrix $A-xI$.*

Proof : Let $m(x)$ be the minimal polynomial of a square matrix A of order n, over a field K. Let $d_i(x)$ and $e_i(x)$ denote respectively the ith determinant divisor and the ith invariant factor of the characteristic matrix $A-xI$, for $i=1,2,\ldots,n$. Then

$$\det(A-xI)=d_n(x)=d_{n-1}(x).\ e_n(x).$$

As $\det(A-xI)$ is a non-null polynomial, $e_n(x)$ is also so.

Let $P(x)$ be the adjoint matrix of $A-xI$. Then $d_{n-1}(x)$ is the h.c f. of the elements of the matrix $P(x)$ in $K[x]$, so that
$$P(x)=d_{n-1}(x).\ D(x),$$
where the h.c.f. of the elements of $D(x)$ is a unit in $K[x]$,

Again, $$P(x).\ (A-xI)=\det(A-xI).\ I,$$
i.e., $$d_{n-1}(x).\ D(x).\ (A-xI)=d_{n-1}(x).\ e_n(x).\ I,$$
i.e., $$D(x).(A-xI)=e_n(x).\ I. \qquad \ldots (1)$$

Hence, by VIII.1.3, A is a root of $e_n(x)$. Then, by VIII.3.1,
$$e_n(x)=h(x).m(x). \qquad \ldots (2)$$

Again, by VIII.1.3,
$$m(x).I=m(xI)=Q(x).\ (A-xI).$$

Then, by (1) and (2),
$$D(x).\ (A-xI)=e_n(x).I=h(x).\ m(x).\ I=h(x).\ Q(x).\ (A-xI).$$

Hence, by the property of the uniqueness of the quotients in division algorithm (vide VIII.1.1), one obtains
$$D(x)=h(x).\ Q(x).$$

But, as the h.c.f. of the elements of $D(x)$ is a unit, $h(x)$ must be a unit in $K[x]$. Hence, by (2), $m(x)$ is an associate of $e_n(x)$, and one may take $m(x)=e_n(x)$.

Since $\psi_A(x)=\det(A-xI)=d_{n-1}(x).e_n(x)=d_{n-1}(x).m(x)$, it follows that :

Corollary 1 : The characteristic polynomial of a square matrix A of order n, over a field K, is its minimal polynomial if, *and only if, the* $(n-1)$*th determinant divisor of the characteristic matrix* $A-xI$ *is a unit in* $K[x]$.

In particular, for the companion matrix $C(f)$, corresponding to the polynomial $f(x)=x_n+b_{n-1}x^{n-1}+ \dots +b_1x+b_0$, defined in VIII.2.2, the following minor of order $n-1$ of the matrix $C(f)-xI$, obtained by omitting the first column and the last row,

$$\begin{vmatrix} 1 & 0 & \dots & 0 & 0 \\ -x & 1 & \dots & 0 & 0 \\ \dots & & \dots & \dots & \\ 0 & 0 & \dots & -x & 1 \end{vmatrix} = 1.$$

Hence, $d_{n-1}(x)=$ h.c.f. of all minors of order $n-1$ of $A-xI=1$.

Corollary 2 : The characteristic polynomial of the companion matrix $C(f)$ *is its minimal polynomial, and is given by* $(-1)^n f(x)$.

A square matrix is said to be *non-derogatory*, if its minimal polynomial is identical with the charateristic polynomial; otherwise, it is said to be *derogatory*. The corollary 1 above gives a necessary and sufficient condition for a matrix to be non-derogatory. In particular, the companion matrix $C(f)$ is non-derogatory. Another example is given by :

Corollary 3 : The Jordan block $J(\lambda)$ *of order n, given below, is non-derogatory,*

$$J(\lambda)=\begin{bmatrix} \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \\ 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda \end{bmatrix}.$$

Proof : The minor of order $n-1$, obtained by omitting the first column and the last row in the characteristic matrix $J(\lambda)-xI$, has the value 1. Hence the $(n-1)$th determinant divisor of $J(\lambda)-xI$ is 1.

In fact, the characteristic polynomial of $J(\lambda)$, which is also its minimal polynomial, is given by $(\lambda-x)^n$.

EXAMPLE 1 : To find the minimal polynomial of the matrix

$$A=\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}:$$

Soln : The characteristic matrix of A is given by

$$A - xI = \begin{bmatrix} 1-x & 1 & 0 \\ 0 & 1-x & 0 \\ 0 & 1 & 1-x \end{bmatrix}.$$

The determinant divisors of $A - xI$ are found to be

$$d_1(x) = 1,\ d_2(x) = 1-x,\ d_3(x) = (1-x)^3.$$

Hence the last invariant factor is

$$e_3(x) = d_3(x)/d_2(x) = (1-x)^2 = x^2 - 2x + 1.$$

It is the required minimal polynomial.

Note : The minimal polynomial of the matrix A is *not irreducible* over the ground field.

Exercise

1. Find the minimal polynomials of the matrices given in exercise 1, page 157, and exercises 1 and 3, page 136.

4. Eigen values of a matrix.

The roots of the characteristic polynomial of a matrix A are called the *eigen values, proper values, characteristic values* or *characteristic roots* of the matrix A. The eigen values do not necessarily lie in the ground field K, but in a suitable extension M of K (viz. M may be taken to be the splitting field of the characteristic polynomial $\psi_A(x)$ over K, in particular, or the algebraic closure of K in general). For a real matrix, the eigen values are, in general, complex numbers. If, however, K is an algebraically closed field, then all the eigen values of any square matrix over K lie in K. Thus for any complex matrix, the eigen values are complex numbers. In particular,

VIII.4.1. *The eigen values of a diagonal matrix, $D = (d_i)$, are given by the diagonal elements $d_1, \ldots, d_n$.*

Proof : The characteristic polynomial of D is given by

$$\psi_D(x) = \det(D - xI) = (d_1 - x)(d_2 - x) \ldots (d_n - x).$$

Hence the eigen values of D are its diagonal elements $d_1, d_2, \ldots, d_n$.

The roots of the characteristic polynomial of a matrix may not be all distinct. If λ be a root of multiplicity r of $\psi_A(x)$, then λ is called an *r-fold* eigen value of A.

VIII.4.2. *If λ be an r-fold eigen value of a matrix A, then 0 is an r-fold eigen value of the matrix $A-\lambda I$.*

Proof: By hypothesis, $\psi_A(x)=(x-\lambda)^r . f(x)$, where $f(\lambda)\neq 0$. Hence

$$\det[(A-\lambda I)-xI]=\det[A-(x+\lambda)I]=\psi_A(x+\lambda)$$
$$=[(x+\lambda)-\lambda]^r . f(x+\lambda)=x^r . f(x+\lambda).$$

For $x=0$, $f(x+\lambda)=f(\lambda)\neq 0$. Hence, it follows that 0 is a root of multiplicity r of $\det[(A-\lambda I)-xI]$; in other words, 0 is an r-fold eigen value of the matrix $A-\lambda I$.

As an important application of the above proposition, one obtains :

VIII.4.3. *If λ be an r-fold eigen value of a matrix A of order n, then the rank of the matrix $A-\lambda I$ is not less than $n-r$.*

Proof: The coefficient of x^r in the characteristic polynomial $\det[(A-\lambda I)-xI]$ of the matrix $A-\lambda I$ is $(-1)^r$.(sum of the principal minors of order $n-r$ of the matrix $A-\lambda I$). As λ is an r-fold eigen value of A, 0 is an r-fold eigen value of the matrix $A-\lambda I$ (by VIII.4.2). Hence, the coefficients of x^{r-1}, $x^{r-2}, \ldots, x$, and the term independent of x in the characteristic polynomial of the matrix $A-\lambda I$ are all null, but the coefficient of x^r is not null. Thus the sum of the principal minors of order $n-r$ of the matrix $A-\lambda I$ is non-null. Consequently, one at least of the minors of order $n-r$ of $A-\lambda I$ must be non-null. Hence the rank of $A-\lambda I$ is not less than $n-r$.

EXAMPLE 1 : For the matrix A, given in example 1, page 160, the characteristic polynomial is $\det(A-xI)=(1-x)^3$. Thus A has only one distinct eigen value, viz. 1 ; and it is a 3-fold eigen value of A. The rank of the matrix $A-1\,I$ is found to be 1.

EXAMPLE 2 : The characteristic polynomial of the matrix,

$$A = \begin{pmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{pmatrix},$$

is $(1-x)^2(5-x)$. Hence 1 and 5 are the eigen values of A. The rank of the matrix $A-1.I$, corresponding to the 2-fold eigen value 1, is found to be 1.

In the above examples it is found that, for an r-fold eigen value λ of an $n \times n$ matrix A, the rank of $A-\lambda I$ is greater than $n-r$ in example 1, and equal to $n-r$ in example 2.

VIII.4.4. *If $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigen values (not necessarily distinct) of a matrix A of order n, over a field K, then $f(\lambda_1)$, $f(\lambda_2), \ldots, f(\lambda_n)$ are the eigen values of the matrix $f(A)$, where $f(x)$ is any polynomial over K.*

Proof: Let Γ be the algebraic closure of K. In $\Gamma[x]$, the characteristic polynomial of A splits up into linear factors,

$$\psi_A(x) = \det(A - xI) = (\lambda_1 - x)(\lambda_2 - x)\ldots(\lambda_n - x).$$

Let $f(x)$ be a polynomial of degree r over K. Then, for any fixed element $\mu \in \Gamma$, $f(x)-\mu$ can be expressed as a product of linear factors in $\Gamma(x)$,

$$f(x) - \mu = a(x-\alpha_1)(x-\alpha_2)\ldots(x-\alpha_r).$$

Correspondingly, one obtains

$$f(A) - \mu I = a(A-\alpha_1 I)(A-\alpha_2 I)\ldots(A-\alpha_r I).$$

$$\begin{aligned}
\text{Then } \det[f(A)-\mu I] &= a^n.\det(A-\alpha_1 I).\det(A-\alpha_2 I)\ldots\det(A-\alpha_r I) \\
&= a^n \psi_A(\alpha_1).\psi_A(\alpha_2)\ldots\psi_A(\alpha_r) \\
&= a^n[(\lambda_1-\alpha_1)(\lambda_2-\alpha_1)\ldots(\lambda_n-\alpha_1)]. \\
&\quad [(\lambda_1-\alpha_2)(\lambda_2-\alpha_2)\ldots(\lambda_n-\alpha_2)]. \\
&\quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\
&\quad [(\lambda_1-\alpha_r)(\lambda_2-\alpha_r)\ (\lambda_n-\alpha_r)] \\
&= [f(\lambda_1)-\mu][f(\lambda_2)-\mu]\ldots[f(\lambda_n)-\mu].
\end{aligned}$$

As the above relation holds for each value of μ, it follows that $\det[f(A)-xI] = [f(\lambda_1)-x][f(\lambda_2)-x]\ldots[f(\lambda_n)-x]$.

Hence $f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_n)$ are the eigen values of $f(A)$.

VIII.4.5. *Every eigen value of a matrix A is also a root of the minimal polynomial of A.*

Proof: Let A be a matrix of order n, over a field K; and let $m(x)$ be its minimal polynomial. Let $e_1(x), e_2(x), \ldots, e_n(x)$ be the invariant factors of the characteristic matrix $A - xI$. Then, by VIII.3.3, $m(x) = e_n(x)$. Also, the characteristic polynomial

$$\psi_A(x) = \det(A - xI) = e_1(x).\ e_2(x)\ .\ e_n(x).$$

Hence any eigen value of A, i.e., any root of $\psi_A(x)$ is necessarily a root of some $e_i(x)$. But (by the property of invariant factors) each $e_i(x)$ is a factor of $e_n(x)$ in $K[x]$. Hence every eigen value of A is necessarily a root of $e_n(x)$, i.e., a root of the minimal polynomial $m(x)$.

Since the minimal polynomial is a factor of the characteristic polynomial, it follows that, if λ be an r-fold eigen value of A, then its multiplicity as a root of the minimal polynomial of A is less than or equal to r. It, however, follows immediately from VIII.4.5 that :

VIII.4.6. *If the eigen values of a matrix A are all distinct (i.e., each is an 1-fold eigen value), then the characteristic ploynomial of A is its minimal polynomial.*

Exercises

1. Find the eigen values of the matrix A, given below :

$$\begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 3 & 0 \\ 2 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & -1 & -1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1+i & 0 & 3i \\ 0 & 2-i & 0 \\ -2 & 0 & 1-i \end{pmatrix}.$$

2. For each eigen value λ of A, in exercise 1, find the rank of the matrix $A - \lambda I$.

3 Show that the eigen values of an upper triangular matrix (i.e., a square matrix (a_{ij}), for which $a_{ij} = 0$ if $i > j$) are given by its diagonal elements (and similarly for a lower triangular matrix).

4. Show that the matrices A and A^t have the same eigen values.

5. Show that, for a complex matrix A, the eigen values of the conjugate matrix A^* are the conjugates of the eigen values of A.

6. If $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigen values of a matrix A (of order n), then, for any scalar u,

(i) $u\lambda_1, u\lambda_2, \dots, u\lambda_n$ are the eigen values of uA ;

(ii) $\lambda_1 - u, \lambda_2 - u, \dots, \lambda_n - u$ are the eigen values of $A - uI$;

(iii) $\lambda_1^m, \lambda_2^m, \dots, \lambda_n^m$ are the eigen values of A^m, where m is any positive integer ;

(iv) $\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_n^{-1}$ are the eigen values of A^{-1}, provided that A is non-singular.

7. Show that, if the rank of a matrix A of order n be r, then 0 is a t-fold eigen value of A where $t \geqslant n - r$.

8. If A and B be two matrices of order n, of which A is non-singular, then

(i) $A^{-1}B$ and BA^{-1} have the same eigen values ;

(ii) B and ABA^{-1} have the same eigen values.

9. Show that, if λ be an eigen value of a non-singular matrix A, then $\lambda^{-1}.\det(A)$ is an eigen value of the adjoint matrix adj(A).

5. Eigen vectors of a matrix.

Let A be a matrix of order n, over a field K, and let $V_n(K)$ denote the linear space formed by the n-tuples $(x_1, \dots, x_n)$, with the co-ordinates x_i from K.

Let $X = (x_1, \dots, x_n)$ be a non-null vector, belonging to $V_n(K)$, such that

$$XA = \lambda X \text{ holds, for some scalar } \lambda \in K ; \qquad \dots (1)$$

then X is called an *eigen vector, proper vector, characteristic vector*, or *invariant vector of the matrix* A. The vector X is then a non-null solution of the equation

$$X(A - \lambda I) = \theta \text{ (the null vector).} \qquad \dots (2)$$

The matrix equation (2) represents a system of linear homogeneous equations in n unknowns $x_1, \dots, x_n$, which has a non-null solution if, and only if,

$$\text{rank } (A - \lambda I) < n, \text{ i.e., } \det(A - \lambda I) = 0,$$

i.e., λ is an eigen value of the matrix A.

Thus one obtains the following proposition :

VIII.5.1. *There exists an eigen vector X of a matrix A, with λ as the corresponding scalar (i e., such that $XA=\lambda X$ holds) if, and only if, λ is an eigen value of A.*

The correspondence between the eigen values and the eigen vectors of a matrix are brought out in the following three propositions.

VIII.5.2. *To each eigen vector X of a matrix A, there corresponds a unique eigen value λ.*

Proof: If possible, let λ_1, λ_2 be two distinct eigen values corresponding to an eigen vector X. Then, from $XA=\lambda_1 X$ and $XA=\lambda_2 X$, one obtains $(\lambda_1-\lambda_2)X=\theta$. And, since X is a non-null vector, it follows that $\lambda_1=\lambda_2$ which is a contradiction.

VIII.5.3. *Let A be a matrix of order n, over a field K. Then*

(i) *to each eigen value λ of A, there corresponds at least one eigen vector X ;*

(ii) *the eigen vectors X, corresponding to a particular eigen value λ, taken together with the null vector θ, form a non-null sub-space of $V_n(K)$ (it is called the characteristic sub-space corresponding to λ) ;*

(iii) *the rank of the characteristic sub-space, corresponding to an r-fold eign value λ, is not greate than rr, and not less than* 1.

Proof: (i) This result follows immediately from VIII.5.1.

(ii) The eigen vectors X, which correspond to an eigen value λ, are given by the non-null solutions of the equation (2). The solutions of (2) are the solutions of a system of linear homogeneous equations, with $A-\lambda I$ as the coefficient matrix ; and such solutions form a sub-space of $V_n(K)$ (vide V 1 1).

(iii) The characteristic sub-space, corresponding to an eigen value λ, consists of the solutions of the equation (2) ; it is the null space of $A-\lambda I$. Hence, by V.1.2, the rank of the characteristic sub-space $=$ nullity of $A-\lambda I=n-\text{rank}(A-\lambda I)$ $\leq n-(n-r)=r$, since for an r-fold eigen value λ, $\text{rank}(A-\lambda I)$

$\geqslant n-r$ (by VIII.4.3). Since there exists at least one eigen vector (which is a non-null vector), the rank of the characteristic subspace $\geqslant 1$.

For an r-fold eigen value λ, r is sometimes called the *algebraic multiplicity of* λ, while the rank of the characteristic subspace corresponding to λ is called the *geometric multiplicity of* λ. It follows from VIII.5.3(iii) that, for any eigen value λ, $1 \leqslant$ *the geometric multiplicity* $\leqslant$ *the algebraic multiplicity.*

VIII.5.4 *If $Y_1,\ldots,Y_m$ be m eigen vectors of an $n\times n$ matrix A, over a field K, such that their corresponding eigen values $\lambda_1,\ldots,\lambda_m$ are distinct, then $Y_1,\ldots,Y_m$ are linearly independent in $V_n(K)$.*

Proof: If possible let $Y_1,\ldots,Y_m$ be linearly dependent over K. Then there exists a linearly dependent subset of vectors Y_i, say r in number, such that any $r-1$ of these vectors form a linearly independent set. Without any loss of generality, let $Y_1,\ldots,Y_r$ be the subset with this property. Then there exist non-null scalars $a_1,\ldots,a_r \in K$, such that

$$a_1Y_1+\ldots+a_rY_r=\theta. \qquad \ldots (3)$$

Then $$(a_1Y_1+\ldots+a_rY_r).\ A=\theta.\ A,$$

i e., $$a_1(Y_1A)+\ldots+a_r(Y_rA)=\theta,$$

or, $$a_1.\lambda_1Y_1+\ldots+a_r.\lambda_rY_r=\theta, \text{ (since } Y_iA=\lambda_iY_i). \qquad \ldots (4)$$

Now, subtracting λ_1-times (3) from (4), one gets

$$a_2(\lambda_2-\lambda_1).\ Y_2+\ldots+a_r(\lambda_r-\lambda_1).Y_r=\theta,$$

which shows that the $r-1$ vectors $Y_2,\ldots,Y_r$ are linearly dependent, since $a_i \neq 0$ and $\lambda_i-\lambda_1 \neq 0$, for $i=2,\ldots,r$. This contradicts the assumption that any $r-1$ vectors in the subset $Y_1,\ldots,Y_r$ are always linearly independent. Hence the given eigen vectors $Y_1,\ldots,Y_m$ form a linearly indepedent set in $V_n(K)$.

It has been pointed out, in discussions on eigen values, that the eigen values of a matrix A of order n, over a field K, need not lie in K. In fact, it may so happen that K may not contain any eigen value of A. In that case, one may look for eigen values of A in a suitable extension M of K. The extension M may, in general, be taken to be the algebraic closure of K, or, in particular, the splitting field of the characteristic polynomial $\psi_A(x)$. Then the eigen vectors lie in $V_n(M)$. Some particular cases will now be considered.

Complex matrices. For any matrix A of order n, over the complex field C, all the n eigen values of A lie in C (as C is algebraically closed), and all the eigen vectors of A lie in V_n (C). One obtains, in particular :

VIII.5.5. (i) *The eigen values of a Hermitian matrix are all real.*

(ii) *The eigen values of a skew-Hermitian matrix are purely imaginary mumbers, or zero.*

(iii) *The eigen values of a unitary matrix have unit modulus.*

Proof: Let λ be an eigen value of a complex matrix A, and let X be an eigen vector corresponding to λ, so that

$$XA = \lambda X. \qquad \dots (5)$$

Then
$$A^{\circ}X^{\circ} = \lambda^{*}X^{\circ}, \qquad \dots (6)$$

where A° denotes the transposed of the conjugate matrix A^{*} of A.

(i) Let A be a Hermitian matrix then $A^{\circ} = A$.

Then, by (5 and (6), one has

$$XAX^{\circ} = (XA)X^{\circ} = (\lambda X)X^{\circ} = \lambda(XX^{\circ}),$$

and
$$XAX^{\circ} = XA^{\circ}X^{\circ} = X(A^{\circ}X^{\circ}) = X(\lambda^{*}X^{\circ}) = \lambda^{*}(XX^{\circ}).$$

Hence $\lambda(XX^{\circ}) = \lambda^{*}(XX^{\circ})$, so that $\lambda = \lambda^{*}$, since $XX^{\circ} \neq 0$.

The eigen value λ is therefore a real number.

(ii) Let A be a skew-Hermitian matrix, then $A^{\circ} = -A$. Then, one obtains, as in (i),

$$XAX^{\circ} = (\lambda X)X^{\circ} = \lambda(XX^{\circ}),$$

and
$$XAX^{\circ} = X(-A^{\circ})X^{\circ} = -X(A^{\circ}X^{\circ}) = -X(\lambda^{*}X^{\circ})$$
$$= -\lambda^{*}(XX^{\circ}),$$

so that $\lambda = -\lambda^{*}$ (since $XX^{\circ} \neq 0$). Hence λ is a purely imaginary number, or zero.

(iii) Let A be a unitary matrix, then $AA^{\circ}=I$. Now, by (5) and (6), one obtains,

$$XX^{\circ}=XIX^{\circ}=X(AA^{\circ}X^{\circ})=(XA).(A^{\circ}X^{\circ})=(\lambda X)(\lambda^{*}X^{\circ})$$
$$=\lambda\lambda^{*}(XX^{\circ}).$$

As $XX^{\circ}\neq 0$, if follows that $\lambda\lambda^{*}=1$, i.e., $|\lambda|=1$.

It should be noted that although the eigen values of a Hermitian matrix A are all real, its eigen vectors do *not* necessarily have real co-ordinates only, since the coefficient matrix $A-\lambda I$ of the equation (2) have complex elements (even with λ as a real number).

Let A be a normal matrix, then $AA^{\circ}=A^{\circ}A$. (where $A^{\circ}=(A^{*})^{t}=(A^{t})^{*}$).

VIII.5.6. *If X be an eigen vector, corresponding to an eigen value λ of a normal matrix A, then X is an eigen vector of the matrix A°, corresponding to the eigen value λ^{*}.*

Proof: Let $B=A-\lambda I$, then

$$BB^{\circ}=(A-\lambda I).(A-\lambda I)^{\circ}=(A-\lambda I)(A^{\circ}-\lambda^{*}I)$$
$$=AA^{\circ}-\lambda A^{\circ}-\lambda^{*}A+\lambda\lambda^{*}I=A^{\circ}A-\lambda^{*}A-\lambda A^{\circ}+\lambda\lambda^{*}I$$
$$=(A^{\circ}-\lambda^{*}I)(A-\lambda I)=(A-\lambda I)^{\circ}(A-\lambda I)=B^{\circ}B.$$

Thus B is also a normal matrix. Again,

$$(XB)(XB)^{\circ}=X(BB^{\circ})X^{\circ}=X(B^{\circ}B)X^{\circ}=(XB^{\circ})(XB^{\circ})^{\circ}. \quad \dots (7)$$

But $XB=X(A-\lambda I)=\theta$, since X is an eigen vector of A corresponding to the eigen value λ. Hence, by (7),

$$XB^{\circ}=\theta, \text{ i.e., } X(A-\lambda I)^{\circ}=\theta$$

or,

$$X(A^{\circ}-\lambda^{*}I)=\theta, \text{ i.e., } XA^{\circ}=\lambda^{*}X.$$

Thus X is an eigen vector of A°, corresponding to λ^{*}.

VIII.5.7. *For a normal matrix, the eigen vectors, corresponding to distinct eigen values, are orthogonal.*

Proof: Let X and Y be two eigen vectors of a normal matrix A, corresponding to the two distinct eigen values λ and μ respectively, so that

$$XA=\lambda X, \text{ and } YA=\mu Y.$$

Then by VIII.5.6, it follows that

$$XA^{\circ} = \lambda^{*}X, \text{ and } YA^{\circ} = \mu^{*}Y.$$

Now, $YA^{\circ}X^{\circ} = Y(XA)^{\circ} = Y(\lambda X)^{\circ} = Y(\lambda^{*}X^{\circ}) = \lambda^{*}(YX^{\circ})$,

and $YA^{\circ}X^{\circ} = (YA^{\circ})X = (\mu^{*}Y)X^{\circ} = \mu^{*}(YX^{\circ})$.

Thus $\lambda^{*}(YX^{\circ}) = \mu^{*}(YX^{\circ})$, i.e., $(\lambda^{*} - \mu^{*})YX^{\circ} = 0$.

Hence $YX^{\circ} = 0$, since $\lambda^{*} \neq \mu^{*}$. Thus the vectors X and Y are orthogonal.

Since every Hermitian, skew-Hermitian or unitary matrix is a normal matrix, it follows that :

Corollary : *For any Hermitian, skew-Hermitian, or unitary matrix, the eigen vectors, corresponding to distinct eigen values, are orthogonal.*

Real matrices. Let A be a matrix of order n, over the field of real numbers **E**. Then the characteristic polynomial of A has real coefficients. Its roots, i.e., the eigen values of A, are in general complex numbers ; some of them may however be real. If λ be an eigen value, which is a real number, then the elements of the matrix $A - \lambda I$ are all real ; and consequently, the corresponding eigen vectors have all the co-ordinates real, i.e., they lie in $V_n(\mathbf{E})$. But, if λ be not a real number, then the elements of $A - \lambda I$ are not all real, and the corresponding eigen vector cannot have all their co-ordinates real ; that is, the eigen vectors do not lie in $V_n(\mathbf{E})$.

As every real symmetric matrix is Hermitian, and every real skew-symmetric matrix is skew-Hermitian, it follows, by VIII.5.5(i)-(ii), that :

VIII.5.8. *The eigen values of a real symmetric matrix are real numbers, and the eigen vectors have real co-ordinates.*

VIII.5.9. *The eigen values of a real skew-symmetric matrix are purely imaginary numbers or zero.*

Thus 0 *is the only possible real eigen value of a real skew-symmetric matrix ; and the corresponding eigen vectors have real co-ordinates.*

As every real orthogonal matrix is unitary, it follows, by VIII.5.5(iii), that :

VIII.5.10. *The eigen values of a real orthogonal matrix have unit modulus.*

Thus 1 *and* −1 *are the only possible real eigen values of a real orthogonal matrix; and their corresponding eigen vectors have real co-ordinates.*

Again, since a real symmetric, skew-symmetric or orthogonal matrix is necessarily a normal matrix, it follows, by VIII 5.7, that :

VIII.5.11. *For a real symmetric, skew-symmetric or orthogonal matrix, the eigen vectors, corresponding to distinct eigen values, are orthogonal.*

EXAMPLE 1 : To find the eigen values and the eigen vectors of the following real matrices :

(i) $A=\begin{pmatrix} 3 & 2 & 2 \\ 1 & 4 & 1 \\ -2 & -4 & -1 \end{pmatrix}$, (ii) $B=\begin{pmatrix} 3 & 1 & 1 \\ 4 & 3 & 1 \\ 0 & 0 & 1 \end{pmatrix}$,

(iii) $G=\begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 0 & 0 & 1 \end{pmatrix}$, (iv) $H=\begin{pmatrix} 1 & -1 & -1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$.

Soln. (i) The characteristic polynomial of A is

$$\det(A-xI)=\begin{vmatrix} 3-x & 2 & 2 \\ 1 & 4-x & 1 \\ -2 & -4 & -1-x \end{vmatrix} = -x^3+6x^2-11x+6$$

$$= -(x-1)(x-2)(x-3).$$

Thus A has three distinct eigen values, $\lambda=1, 2, 3$.

The eigen vectors $X=(x_1, x_2, x_3)$, corresponding to an eigen value λ, are given by the non-null solution vectors of the equation :

$$X(A-\lambda I)=(x_1, x_2, x_3)\begin{pmatrix} 3-\lambda & 2 & 2 \\ 1 & 4-\lambda & 1 \\ -2 & -4 & -1-\lambda \end{pmatrix}=(0, 0, 0),$$

i.e.,

$$\left.\begin{aligned} (3-\lambda)x_1 + x_2 - 2x_3 &= 0 \\ 2x_1 + (4-\lambda)x_2 - 4x_3 &= 0 \\ 2x_1 + x_2 - (1+\lambda)x_3 &= 0 \end{aligned}\right\} \quad \ldots \ (8)$$

For $\lambda=1$, the equations (8) reduce to

$$\left.\begin{aligned} 2x_1 + x_2 - 2x_3 &= 0 \\ 2x_1 + 2x_2 - 4x_3 &= 0 \\ 2x_1 + x_2 - 2x_3 &= 0 \end{aligned}\right\},$$

and the corresponding eigen vectors are given by $r(1, 2, 2)$, where r assumes all non-zero real values.

For $\lambda=2$, the equations (8) reduce to

$$\left.\begin{array}{r} x_1+x_2-2x_3=0 \\ 2x_1+2x_2-4x_3=0 \\ 2x_1+x_2-3x_3=0 \end{array}\right\},$$

and the corresponding eigen vectors are given by $r(1,1,1)$. where r assumes all non-zero real values.

For $\lambda=3$, the equations (8) reduce to

$$\left.\begin{array}{r} x_2-2x_3=0 \\ 2x_1+x_2-4x_3=0 \\ 2x_1+x_2-4x_3=0 \end{array}\right\},$$

and the corresponding eigen vectors are given by $r(1,2,1)$, where r assumes all non zero real values.

(ii) The characteristic polynomial of the matrix B is

$$-x^3+7x^2-11x+5=-(x-1)^2(x-5).$$

Thus B has only two distinct eigen values, $\lambda=1$ and 5.

The eigen vectors $X=(x_1,x_2,x_3)$, corresponding to an eigen value λ, are given by the non-null solution vectors of the equation :

$$(x_1,x_2,x_3)\begin{pmatrix} 3-\lambda & 1 & 1 \\ 4 & 3-\lambda & 1 \\ 0 & 0 & 1-\lambda \end{pmatrix}=(0,0,0),$$

i.e.,

$$\left.\begin{array}{lllr} (3-\lambda)x_1 & +4x_2 & & =0 \\ x_1 & +(3-\lambda)x_2 & & =0 \\ x_1 & +x_2 & +(1-\lambda)x_3 & =0 \end{array}\right\}. \qquad \ldots (9)$$

For $\lambda=1$, the equations (9) reduce to

$$\left.\begin{array}{r} 2x_1+4x_2=0 \\ x_1+2x_2=0 \\ x_1+x_2=0 \end{array}\right\},$$

and the corresponding eigen vectors are given by $r(0,0,1)$, where r assumes all non zero real values.

For $\lambda=5$, the equations (9) reduce to

$$\left.\begin{array}{lr} -2x_1+4x_2 & =0 \\ x_1-2x_2 & =0 \\ x_1+x_2-4x_3 & =0 \end{array}\right\},$$

and the corresponding eigen vectors are given by r (8, 4, 3), where r assumes all non-zero real values.

(iii) The characteristic polyonomial of the matrix G is

$$-x^3+7x^2-11x+5=-(x-1)^2(x-5).$$

Thus G has only two distinct eigen values, $\lambda=1$ and 5.

The eigen vectors, corresponding to an eigen value λ, are given by the non-null solution vectors of the equations :

$$\left.\begin{array}{lllr} (3-\lambda)x_1 & +2x_2 & & =0 \\ 2x_1 & +(3-\lambda)x_2 & & =0 \\ x_1 & +x_2 & +(1-\lambda)x_3 & =0 \end{array}\right\}. \qquad \ldots (10)$$

For $\lambda=1$, the equations (10) reduce to

$$\left.\begin{array}{r} 2x_1+2x_2=0 \\ 2x_1+2x_2=0 \\ x_1+x_2=0 \end{array}\right\},$$

and the solution vectors form a sub-space of rank 2, for which $(-1,1,0)$ and $(0,0,1)$ form a basis. The non-null vectors of this sub-space are the eigen vectors corresponding to the eigen value 1.

For $\lambda=5$, the equations (10) reduce to

$$\left.\begin{array}{rl} -2x_1+2x_2 & =0 \\ 2x_1-2x_3 & =0 \\ x_1+x_2-4x_3 & =0 \end{array}\right\}.$$

and the corresponding eigen vectors are given by $r(2,2,1)$, where r assumes all non-zero real values.

(iv) The characteristic polynomial of the matrix H is

$$-x^3-x^2-x-1=-(x+1)\,(x-i)\,x+i).$$

Thus H has three distinct eigen values $\lambda=-1, i, -i$. Only one of the eigen values is real.

The eigen vectors, corresponding to an eigen value λ, are given by the non-null solution vectors of the equations :

$$\left.\begin{array}{rrrl} (1-\lambda)x_1 & +\;x_2 & +\;x_3 & =0 \\ -x_1 & -\;(1+\lambda)x_2 & & =0 \\ -x_1 & & -(1+\lambda)x_3 & =0 \end{array}\right\}, \quad \dots (11)$$

For $\lambda=-1$, the equations (11) reduce to

$$\left.\begin{array}{rl} 2x_1+x_2+x_3 & =0 \\ -x_1 & =0 \\ -x_1 & =0 \end{array}\right\},$$

and the corresponding eigen vectors are given by $r(0,1,-1)$, where r assumes all non zero real values.

For the eigen values i and $-i$, the eigen vectors lie in $V_3(C)$, and their co-ordinates can not all be real.

For $\lambda=i$, the equations (11) reduce to

$$\left.\begin{array}{rrrl} (1-i)x_1+ & x_2+ & x_3 & =0 \\ x_1 & +\;(1+i)x_2+ & & =0 \\ x_1 & & +(1+i)x_3 & =0 \end{array}\right\},$$

and the corresponding eigen vectors are given by $\rho(1+i,-1,-1)$, where ρ assumes all non-zero complex values.

For $\lambda=-i$, the equations (11) reduce to

$$\left.\begin{array}{rrrl} (1+i)x_1+ & x_2+ & x_3 & =0 \\ x_1 & +\;(1-i)x_2 & & =0 \\ x_1 & & +(1-i)x_3 & =0 \end{array}\right\},$$

and the corresponding eigen vectors are given by $\rho(1-i,-1,-1)$, where ρ assumes all non-zero complex values.

Note : Both the matrices B and G have the same characteristic polynomial. The rank of the invariant sub-space corresponding to the 2-fold eigen value 1 is 1 for the matrix B, while it is 2 in case of the matrix G.

Exercises

1. For the following matrices, find (*a*) the eigen values and the corresponding eigen vectors, and (*b*) the ranks of the characteristic sub-spaces corresponding to the different eigen values :

(i) $\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{pmatrix}$, (ii) $\begin{pmatrix} 2 & 5 & 0 \\ 0 & 2 & 0 \\ -1 & 3 & 2 \end{pmatrix}$, (iii) $\begin{pmatrix} 1 & -2 & 2 \\ -2 & 1 & 0 \\ 0 & -3 & 4 \end{pmatrix}$,

(iv) $\begin{pmatrix} 2 & 1-i & 2+i \\ 1+i & -3 & 3+4i \\ 1-i & 3-4i & 0 \end{pmatrix}$, (v) $\begin{pmatrix} \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{3\sqrt{2}} & \frac{1}{3\sqrt{2}} & \frac{4}{3\sqrt{2}} \end{pmatrix}$,

(vi) $\begin{pmatrix} 7 & -2 & 1 \\ -2 & 10 & -2 \\ 1 & -2 & 7 \end{pmatrix}$,

(vii) $\begin{pmatrix} 2 & 0 & 0 & 0 \\ 3 & 1 & -4 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 5 & 0 & 1 \end{pmatrix}$, (iiiv) $\begin{pmatrix} 3 & 0 & 0 & 0 \\ -1 & 1 & a & 0 \\ 4 & 0 & 1 & 0 \\ 5 & 2 & 2a+1 & 2 \end{pmatrix}$, for $a=0$, and $a \neq 0$.

2. Prove (independently) that eigen vectors corresponding to distinct eigen values of a unitary, Hermitian, real symmetric, or a real orthogonal matrix are orthogonal.

5.1. Application (Fixed points of a collineation of a Projective space).

A collineation of the n-dimensional projective space P^n, over a field K, is given by an equation

$$\rho X' = XA, \qquad \dots (1)$$

where $X' = (x_1', \dots, x_{n+1}')$, $X = (x_1, \dots, x_{n+1}) \in V_{n+1}(K)$, $A = (a_{ij})$ is a square matrix of order $n+1$, over K, and ρ is an arbitrary non-null element of K. The collineation (1) transforms a point P with the co-ordinate vectors sX to the point P' with the co-ordinate vectors tX' (s and t being arbitrary non-null scalars). Hence, a point P with the co-ordinate vectors sX remains fixed under the collineation (1) if, and only if,

$\rho X = XA$ holds, for some non-null scalar ρ,

that is, if and only if X be an eigen vector of the matrix A. The scalar ρ is then an eigen value of A, and X is an eigen vector corresponding to ρ.

The collineation (1) is said to be singular or non-singular, according as the matrix A is singular or non-singular (i.e., according as $\det(A)=0$ or $\neq 0$). For a singular collineation, the equation $XA=\theta$ has always a non-null solution. As the null vector does not represent any point, the images of some points are non-existent (such points are called singular points). Such points do not exist in the case of a non-singular collineation. Also different points are transformed to different points if, and only if, the collineation is non-singular

It will be assumed, in what follows, *that the collineation* (1) *is non-singular*.

It has been shown that :

(A) The eigen vectors corresponding to an eigen value λ, taken together with the null vector, form a linear sub-space of $V_{n+1}(K)$, called the characteristic sub-space corresponding to λ.

(B) The rank of the characteristic sub-space corresponding to an r-fold eigen value λ is $(n+1) - \text{rank}\ (A-\lambda I)$; and it is less than or equal to r, and is greater than or equal to 1.

It is also known that :

(C) In the projective space P^n, the non-null vectors in a linear sub-space of rank $r+1$ represent all points incident with a projective sub-space P^r of (projective) dimension r. Thus, in particular,

(i) the non null vectors in a linear sub-space of rank 1 represent a point ;

(ii) the non-null vectors in a linear sub-space of rank 2 represent all points lying on a line, that is, a row of points ;

(iii) the non-null vectors in a linear sub-space of rank 3 represent all points lying on a plane, that is, a plane field of points.

(D) Every eigen vector X of the matrix A determines a linear sub-space of rank 1, given by sX (as s runs over K), which remains invariant under the collineation (1). Consequently, an characteristic sub-space of rank $m+1$ (corresponding to some r fold eigen value λ) determines a projective sub-space of dimension m, *every point of which remains fixed under the collineation.*

The fixed points of a given collineation (1) are, therefore, found by :

(a) finding out the distinct eigen values of A, with their respective multiplicities, *lying in the ground field K* ; and then

(b) finding out the characteristic sub-spaces corresponding to the distinct eigen values.

The characteristic sub-spaces thus found, represent projective sub-spaces, which are *point-wise fixed* under the collineation. Also all fixed points lie on these projective sub-spaces.

Since eigen vectors, correspondiog to distinct eigen values, are linearly independent, the projective sub-spaces, containing the fixed points are correspondingly independent in respect of their incidence relations. For instance, if there is a fixed point corresponding to some eigen value, and also a line of fixed points corresponding to another eigen value, the (first) point must not lie on the line of fixed points.

The principle outlined above will now be illustrated by considering the cases of the collineations of a real projective plane, and those of a real projective 3-space.

Fixed points of a non-singular collineation of the real projective plane. In the case of the real projective plane P^2, $n=2$, and the ground field K is the field of real numbers. For a collineation (1) of P^2, A is a matrix of order 3. Hence, the characteristic polynomial $\psi_A(x)$ is a cubic with real coefficients. Among its roots, i.e., the eigen values of A, either (i) all the three are real, or (ii) one is real and the two others are imaginary (forming a conjugate pair).

Firstly, let all the three eigen values of A be real. Then the following cases arise :

(I) There exist three distinct eigen values λ_1, λ_2 and λ_3. Each of λ_1, λ_2 and λ_3 is then an 1-fold eigen value of A ; and for each one of these eigen values, the corresponding characteristic sub-space is of rank 1. Hence, *there exist exactly three distinct* *fixed points.* Also, since $\lambda_1, \lambda_2, \lambda_3$ are distinct, the corresponding eigen vectors are linearly independent. Hence *the three fixed points are non collinear.*

(II) There exist two distinct eigen values λ_1 and λ_2, of which λ_1 is 2-fold, and λ_2 is simple. The characteristic sub space, corresponding to λ_2, is of rank 1; and correspondingly there exists a fixed point, say M. The characteristic sub-space, corresponding to λ_1, has rank $=2$ or 1 [according as rank$(A-\lambda_1 I)=1$ or 2]; correspondingly, there exists either a row of fixed points (to which M does not belong), or only one fixed point distinct from M. Thus, there are two possiblities in this case :

(i) *There exists a row of fixed points and another fixed point not belonging to this row* [when rank $(A-\lambda_1 I)=1$]. Such a collineation is known as a *Homology*, with the base of the row as its *axis*, and the isolated fixed point as its *centre.*

(ii) *There exist exactly two distinct fixed points* [when rank of $(A-\lambda_1 I)=2$].

(III) There exists only one distinct eigen value λ with multiplicity 3. Then λ is a 3-fold eigen value. The corresponding characteristic sub-space has rank $=3$, 2 or 1 [according as rank of $(A-\lambda I)=0$, 1 or 2]. Thus there are three possibilities in this case :

(i) *There exists a plane field of fixed points ; that is, every point of the projective plane* P^2 *remains fixed.* The collineation is, in this case, the *identity transformation.* In fact $(A-\lambda I)=O$ implies that $A=\lambda I$ is a scalar matrix, and the equation (1) of the collineation reduces to $\rho X'=XA=X.\lambda I=\lambda X$, i.e., $\sigma X'=X$.

(ii) *There exists a row of fixed points* [when rank $(A-\lambda I)=1$]. The collineation is, in this case, known as an *Elation.*

(iii) *There exists only one fixed point* [when rank of $(A-\lambda I)=2$].

Next, let there exist only one real eigen value λ, the two other eigen values being imaginary numbers. The characteristic sub-space, corresponding to λ, has rank 1. Thus, in this case :

(IV) *There exists only one fixed point.*

The non singular collineations of the real projective plane can be divided into seven distinct classes, as above. For each of the cases III(iii) and IV, there exists only one fixed point; but the two cases differ in other geometrical properties (vide example 2, page 181).

Fixed points of a non-singular collineation of the real projective 3-space. In this case $n=3$, and A is a matrix of order 4. The characteristic polynomial $\psi_A(x)$ is a biquadratic with real coefficients. Among its roots, i.e., the eigen values of A, (i) either all the four are real, or (ii) only two are real, or (iii) none is real.

Firstly, let all the four eigen values of A be real. Then the following cases arise.

(I) There exist four distinct eigen values $\lambda_1, \lambda_2, \lambda_3$ and λ_4. Each of these being an 1-fold eigen value, the corresponding characteristic sub-space is of rank 1. And correspondingly, *there exist exactly four distinct fixed points, no three of which are collinear.*

(II) There exist three distinct eigen values λ_1, λ_2 and λ_3, of which λ_1 is a 2-fold eigen value, and the two others are simple. Corresponding to the two 1-fold eigen values λ_2 and λ_3, the characteristic sub-spaces are of rank 1; and correspondingly, there exist two distinct fixed points M and N, say. The rank of the characteristic sub-space, corresponding to the 2-fold eigen value λ_1, is 2 or 1 [according as rank$(A-\lambda_1 I) = 2$ or 3]; correspondingly, there exists either a row of fixed points, or only one fixed point H (say). Thus there are two possibilities in this case:

(i) *There exists a row fixed points* [when rank $(A-\lambda_1 I) = 2$], *and two other fixed points M and N.* The points M, N cannot be collinear with any point on the row; hence *the line joining M and N must be skew to the base of the row.*

(ii) *There exist exactly three distinct fixed points M, N and H, which are non-collinear* [when rank $(A-\lambda_1 I) = 3$].

(III) There exist two distinct eigen values λ_1 and λ_2, of which either (α) both λ_1 and λ_2 are 2-fold, or (β) λ_1 is 3-fold and λ_2 is simple. Arguing as above, one obtains the following possible cases :

α(i) *There exist two distinct rows of fixed points, whose bases are skew* [when rank $(A-\lambda_1 I)=2$ and rank $(A-\lambda_2 I)=2$].

α(ii) *There exists a row of fixed points and another fixed point, not belonging to the row* [when of rank $(A-\lambda_1 I)$ and rank $(A-\lambda_2 I)$, one has the value 2, and the other one has the value 3].

α(iii) *There exist exactly two distinct fixed points* [when rank $(A-\lambda_1 I)=3$ and rank $(A-\lambda_2 I)=3$].

β(i) *There exists a plane field of fixed points and another fixed point, outside the plane field* [when rank $(A-\lambda_1 I)=1$]. Such a collineation is known as a *Homology.*

β(ii) *There exists a row of fixed points and another fixed point, not belonging to the row* [when rank $(A-\lambda_1 I)=2$].

β(iii) *There exist exactly two distinct fixed points* [when rank $(A-\lambda_1 I)=3$].

In all the three cases for (β), rank $(A-\lambda_2 I)=3$.

(IV) There exists only one distinct eigen value λ with multiplicity 4. Then λ being a 4-fold eigen value, the rank of the associated characteristic sub-space is 4, 3, 2 or 1. Thus one obtains the following possible cases :

(i) *The fixed points form the real projective 3-space* [when rank $(A-\lambda I)=0$]. That is, every point of the real projective 3-space P^3 remains fixed. The collineation is the *identity transformation.*

(ii) *There exists a plane field of fixed points* [when rank of $(A-\lambda I)=1$]. The collineation is, in this case, called an *Elation.*

(iii) *There exists a row of fixed points* [when rank of $(A-\lambda I)=2$].

(iv) *There exists exactly one fixed point* [when rank of $(A-\lambda I)=3$].

Secondly, let there be only two real eigen values (identical or distinct), the other two being imaginary numbers. The following two cases arise :

(V) The two real eigen values are identical Thus there exists a 2-fold eigen value λ ; then the rank of the characteristic sub-space is 2 or 1. There are two possibilities in this case :

(i) *There exists a row of fixed points* [when rank $(A-\lambda I)=2$].

(ii) *There exists exactly one fixed point* [when rank of $(A-\lambda I)=3$].

(VI) The two real eigen values are distinct. Thus there exist two distinct 1-fold eigen values λ_1 and λ_2. The corresponding invariant sub-spaces have rank 1. Hence *there exist exactly two distinct fixed points.*

Finally, when no eigen value is real, one obtains :

(VII) *There are no fixed points.*

The non-singular collineations of the real projective 3-space can be divided into 17 different classes, as obtained above, in respect of the real eigen values (and their multiplicities) of the matrices of the collineations. As for fixed points, some of these cases have the same behaviour. E.g. the cases III-β(iii) and VI [or IV(iii) and V(i), or IV(iv) and V(ii)] are found to have the same type of fixed points. But the cases are distinct, which can be shown by considering the fixed planes and their incidence relations with the fixed points (vide exercises 3-4, page 182).

EXAMPLE 1 : Let $\rho X'=XM$ be a collineation of the real projective plane, whose matrix M is taken to be one of the matrices A, B, G and H in example 1, page 171. To find the fixed points in each case.

Soln. The eigen values and the corresponding characteristic sub spaces of the matrices A, B, G and H have been found in example 1, page 171. It follows therefrom that :

(i) When $M=A$, there exist three distinct fixed points, which are non-collinear (Case I).

(ii) When $M=B$, there are two distinct fixed points (Case II (ii)).

(iii) When $M=G$, there exists a row of fixed points and an isolated fixed point. The collineation is a Homology (Case II(i)).

(iv) When $M=H$, there exists only one fixed point (Case IV).

EXAMPLE 2 : Consider the two collineations of the real projective plane, whose matrices are given by

$$A=\begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & 2 \end{pmatrix}, \qquad B=\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix},$$

Their characteristic polynomials are

$$\psi_A(x)=(2-x)^3, \text{ and } \psi_B(x)=(2-x)(x^2-4x+5).$$

There exists only one distinct real eigen value, viz. 2, with the multiplicity 3, for the matrix A. Also, rank of $(A-2I)=2$. Thus the collineation $\rho X'=XA$ is of the type III(iii), page 177. Hence there exists only one fixed point, under this collineation.

There exists only one real eigen value, viz. 2, the two other eigen values being imaginary numbers, for the matrix B. Thus the collineation $\rho X'=XB$ is of the type IV, pages 177-78. Hence there exists only one fixed point, under this collineation also.

Thus, under each of the two given collineations, there exists only one fixed point. Dually, there exists only one fixed line, under each of them.

The fixed points of the two collineations are given respectively by the solutions of the matrix equations :

$$(x_1,x_2,x_3)\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}=(0,0,0), \text{ and } (x_1,x_2,x_3)\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}=(0,0,0).$$

Thus the fixed points P and Q of the two collineations $\rho X'=XA$ and $\rho X'=XB$ have respectively the co-ordinates (0, 0, 1) and (1, 0, 0).

The fixed lines, in line co-ordinates (u_1,u_2,u_3), of the two collineations are given respectively by the solutions of the following matrix equations (whose matrices are the transpose of the corresponding matrices for fixed points) :

$$(u_1\ u_2,u_3)\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}=(0,0,0), \text{ and } (u_1,u_2,u_3)\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}=(0,0,0).$$

Thus the fixed lines p and q of the two collineations $\rho X'=XA$ and $\rho X'=XB$ have respectively the co-ordinates (1, 0, 0) and (1, 0, 0).

In the case of the collineation $\rho X'=XA$, its fixed point P lies on the fixed line p. But in the case of the collineation $\rho X'=XB$, its fixed point Q does not lie on the fixed line q.

Thus it is found that although the above two collineations have similar behaviour as to their fixed points, they behave differently in other respects.

Exercises

1. Find the fixed points of the collineations $\rho X' = XA$, where A is given by the non-singular matrices in exercise 1, page 174.

2. Construct non-singular matrices illustrating

(i) the seven different types of collineations of the real projective plane (discussed above);

(ii) the seventeen different types of collineations of the real projective 3-space (discussed above).

3. For each of the cases III-β(iii) and VI, pages 179-80, there exist exactly two distinct fixed points, and (dually) two distinct fixed planes. Show, by constructing suitable examples, that in case III-β(iii), one of the fixed points lies on both the fixed planes, and one of the fixed planes passes through both the fixed points. While, in case VI, each fixed point lies on only one of the fixed planes, and each fixed plane passes through only one of the fixed points.

4. Distinguish, as in exercise 3, between the cases IV(iii) and V(i), and between the cases IV(iv) and V(ii), by constructing suitable examples.

CHAPTER IX

SIMILARITY OF MATRICES

176

1. Similarity of matrices.

Let Λ_n be the set of all square matrices of order n, over a field K A matrix B is said to be *similar* to a matrix A, over K, if there exists a non-singular matrix P, such that $B = PAP^{-1}$, where $A, B, P \in \Lambda_n$. As the non-singular matrices, belonging to Λ_n form a group, it follows readily (as in the case of congruence) that *similarity of matrices in Λ_n is an equivalence relation.* Consequently, Λ_n is partitioned into disjoint classes of similar matrices. To find the normal forms, it is helpful to find some invariants of a given matrix, under similarity.

IX.1.1. *Two similar matrices have the same characteristic polynomial.*

Proof: Let $\psi_A(x)$ and $\psi_B(x)$ be the characteristic polynomials of the matrices A and B, belonging to Λ_n. And let B be similar to A, so that $B = PAP^{-1}$, where P is a non-singular matrix. Then

$$\psi_B(x) = \det(B - xI) = \det(PAP^{-1} - xI)$$

$$= \det(PAP^{-1} - P.xI.P^{-1})$$

(since the scalar matrix xI is commutative with P)

$$= \det[P(A - xI)\,P^{-1}] = \det P.\det(A - xI).\det P^{-1}$$

$$= \det(A - xI)\ \det P.\det P^{-1} = \det(A - xI) = \psi_A(x).$$

Since the eigen values of a matrix are the roots its characteristic polynomial, it follows that :

Corollary: *Two similar matrices have the same eigen values.*

The converse of IX.1.1 and, hence, also that of the Corollary are *not true*, as can be seen from the following example.

EXAMPLE 1: The matrices $G = \begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix}$ and $H = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ have the same characteristic polynomial $(2-x)^2$. But, H being a scalar matrix, $PHP^{-1} = H$, for any matrix P, so that there is no matrix similar to H other than H itself. Consequently, G cannot be similar to H.

A necessary and sufficient condition for matrices to be similar has been obtained below in IX 1.3. The following lemma is useful for that purpose :

IX.1.2. *If* $F(x)=F_1x+F_0$ *and* $G(x)=G_1x+G_0$ *be two matric polynomials over* Λ_n, *of which* $G(x)$ *is proper, then* $G(x)$ *is equivalent to* $F(x)$ *over* $K[x]$ *if, and only if, they are equivalent over the ground field* K.

Proof: It is evident that if $G(x)$ be equivalent to $F(x)$ over K, then they are also so over $K[x]$.

Conversely, let $G(x)$ be equivalent to $F(x)$ over $K[x]$, then $G(x)=S(x)F(x)T(x)$, where $S(x)$ and $T(x)$ are products of unimodular elementary matrices over $K[x]$. Thus

$$S(x)F(x)=G(x)T^{-1}(x) \quad \text{and} \quad F(x)T(x)=S^{-1}(x)G(x). \qquad \ldots (1)$$

By division algorithms, one gets,

$$S(x)=G(x)Q_l+R_l \quad \text{and} \quad T(x)=Q_rG(x)+R_r, \qquad \ldots (2)$$

where R_l, $R_r \in \Lambda_n$. Now, by (1) and (2),

$$\begin{aligned}
G(x)=S(x)\,F(x)\,T(x) &= [G(x)Q_l+R_l]\;F(x)T(x)\\
&=R_lF(x)T(x)+G(x)Q_lS^{-1}(x)G(x)\\
&=R_lF(x)[Q_rG(x)+R_r]+G(x)Q_lS^{-1}(x)G(x)\\
&=R_lF(x)R_r+[S(x)-G(x)Q_l]F(x)Q_rG(x)\\
&\qquad\qquad +G(x)Q_lS^{-1}(x)G(x)\\
&=R_lF(x)R_r+[S(x)F(x)]Q_rG(x)\\
&\qquad\qquad +G(x)[-Q_lF(x)Q_r+Q_lS^{-1}(x)]G(x)\\
&=R_lF(x)R_r+G(x)D(x)G(x),
\end{aligned}$$

where $D(x)=T^{-1}(x)Q_r-Q_lF(x)Q_r+Q_lS^{-1}(x)$.

Thus $$G(x)-R_lG(x)R_r=G(x)D(x)G(x). \qquad \ldots. (3)$$

Now, $D(x)$ must be the null matrix, otherwise the matric polynomial on the right hand side of (3) has degree $\geqslant 2$, while that on the left hand side has degree $\leqslant 1$. Hence $G(x)=R_lF(x)R_r$, so that $G(x)$ is equivalent to $F(x)$ over K.

IX.1.3. *Two square matrices* A, B *of order* n, *over a field* K, *are simlar if, and only if, the characteristic matrices* $A-xI$ *and* $B-xI$ *are equivalent over (the Euclidean domain)* $K[x]$.

Proof: Let A and B be similar, so that $B=PAP^{-1}$, where P is a non-singular matrix. Then

$$P(A-xI)P^{-1}=PAP^{-1}-P.xI.P^{-1}=B-xI,$$

(since the scalar matrix xI is commutative with the matrix P). Hence $A-xI$ and $B-xI$ are equivalent over $K[x]$, since P and P^{-1} are two non-singular matrices over K and hence also over $K[x]$.

Conversely, let $A-xI$ and $B-xI$ be equivalent over $K[x]$. Then they are also equivalent over K by the lemma IX.1.2. Hence, there exist non-singular matrices P, Q, over K, such that

$$B-xI=P(A-xI)Q=PAQ-P.xI.Q=PAQ-xPQ\,;$$

then,

$$B=PAQ, \text{ and } I=PQ.$$

Thus $Q=P^{-1}$, so that $B=PAP^{-1}$. Hence A and B are similar.

It now follows immediately, by VI.3.5 (ii) and VI.3.7, that

Corollary 1: *Two square matrices A, B of order n, over a field K, are similar if, and only if, the characteristic matrices $A-xI$ and $B-xI$ have the same set of invariant factors in $K[x]$.*

Corollary 2: *Two square matrices A, B of order n, over a field K, are similar if, and only if, the characteristic matrices $A-xI$ and $B-xI$ have the same rank and the same set of elementary divisors in $K[x]$.*

Since the minimal polynomial of a matrix A of order n, over a field K, is given by the nth invarant factor of the characteristic matrix $A-xI$ (vide VIII.3.3), it follows from Corollary 1:

IX.1.4. *Two similar matrices have the same minimal polynomial.*

Exercises

1. Two similar matrices have the same determinant, and the same trace.

2. If X be an eigen vector of a square matrix A, corresponding to an eigen value λ, then XP^{-1} is an eigen vector of the matrix PAP^{-1}, corresponding to the same eigen value λ.

3. If $f(x)$ be any polynomial, then $f(PAP^{-1}) = P.f(A).P^{-1}$. Deduce that if a square matrix A be a root of a polynomial, then the matrices similar to A are also roots of the same polynomial.

4. Every square matrix is similar to its transpose.

5. The matrices B and G in example 1, page 171, have the same characterisitic polynomial, but they are not similar.

2. Similarity to a diagonal matrix.

The normal forms of matrices (under equivalence, congruence, Hermitian congruence) had so far been diagonal matrices, except for the case of skew-symmetric matrices under congruence. So, it is perfectly natural to enquire whether the same is also true under similarity. And it will be found that it is *not* alway so.

It has been proved that two similar matrices have the same eigen values (vide Corollary to IX.1.1), and that for a diagonal matrix, the diagonal elements are its eigen values (vide VIII.4.1), Hence it follows that :

IX.2.1. *If a matrix A be similar to a diagonal matrix D, then the diagonal elements of D are the eigen values of A*

Thus, if a matrix A be similar to a diagonal matrix D, then the diagonal elements of D are determined uniquely by A.

To obtain a necessary and sufficient condition for a matrix to be similar to a diagonal matrix, the following two lemmas are found to be useful.

IX 2.2. *For square matrices P, A, D of order n, over a field K, $PA = DP$ holds, where D is a diagonal matrix with the diagonal elements d_i, if, and only if, the ith row-vector of P is an eigen vector of A, having d_i as the corresponding eigen value, for $i = 1, 2, .., n$.*

Proof : Let $(p_{i1}, \ldots p_{in})$ denote the ith row-vector of P. Then

the ith row vector of $PA = (p_{i1}, \ldots, p_{in}). A$,

and the ith row-vector of $DP = d_i.(p_{i1}, \ldots, p_{in})$.

Hence $PA=DP$ if, and only if,

$$(p_{i1},\ldots,p_{in}).A = d_i.(p_{i1},\ldots,p_{in}), \text{ for } i=1,2,\ldots,n\ ;$$

that is, if, and only if, $(p_{i1},\ldots,p_{in})$ is an eigen vector of A, having d_i as the corresponding eigen value, for $i=1,2,\ldots,n$.

It follows immediately from the proposition IX.2.2 that :

IX.2.3. *If a matrix A be similar to a diagonal matrix $D=(d_i)$, then for any non-singular matrix P, such that $PAP^{-1}=D$, the ith row-vector of P is an eigen vector of A, corresponding to the eigen value d_i, for $i=1,2\ldots,n$*

Proof : $PAP^{-1}=D$ if, and only if, $PA=DP$. And the result follows from IX.2.2.

IX 2.4. *A square matrix A of order n, over a field K, is similar to a diagonal matrix if, and only if, the set of eigen vectors of A includes a basis of $V_n(K)$ (i.e., if there exist n eigen vectors of A, which are linearly independent over K).*

Proof : Let A be similar to a diagonal matrix $D=(d_i)$. Then there exists a non-singular matrix P, such that $PAP^{-1}=D$, that is, $PA=DP$. Then the row-vectors of P are eigen vectors of A, by IX.2.2 Also, since P is non-singular, its row-vectors are linearly independent over K. Thus the row vectors of P, n in number, are eigen vectors of A. which are linearly independent over K ; and they form a basis of $V_n(K)$.

Conversely, let $\pi_i\ (p_{i1},\ldots,p_{in})$, for $i=1,2\ldots,n$, be eigen vectors of A, forming a basis of $V_n(K)$; that is, $\pi_1,\ldots,\pi_n$ are n eigen vectors of A, which are linearly independent over K Let d_i be the eigen value of A, corresponding to the eigen vector π_i, for $i=1,2,\ldots,n$. If, now, P be the square matrix, having $\pi_1,\ldots,\pi_n$ as its row-vectors, then it follows by IX.2.2 that $PA=DP$, where D is the diagonal matrix, with $d_1,\ldots d_n$ as its diagonal elements. As the row-vectors of P are linearly independent over K, it follows that P is non-singular. Hence P^{-1} exists, and one obtains $PAP^{-1}=D$. Thus A is similar to the diagonal matrix D.

A matrix, which is similar to a diagonal matrix, is said to be *diagonable*.

An example of a matrix, which is *not diagonable*, is given below :

EXAMPLE 1 : The matrix B in example 1 (ii), page 171, has two distinct eigen values 1 and 5 (of which 1 is 2-fold, and 5 is simple). The eigen vectors, corresponding to the eigen value 1, are given by $r(0, 0, 1)$, and those, corresponding the eigen value 5, are given by $r(8, 4, 8)$, where r assumes all non-zero real values. As it is impossible to choose 3 eigen vectors, which are linearly independent, it follows by IX.2.4 that the matrix B cannot be similar to a diagonal matrix.

One may readily deduce the following *sufficient* condition for a matrix to be diagonable :

IX.2.5. *If the eigen values $\lambda_1, \ldots, \lambda_n$ of a matrix A of order n, over a field K, lie in K and are distinct, then A is similar to the diagonal matrix D, having $\lambda_1, \ldots, \lambda_n$ as its diagonal elements.*

Proof : Let X_i be an eigen vector of A, corresponding to the eigen value λ_i, for $i=1,2,\ldots,n$. Then, by VIII.5.4, the eigen vectors $X_1, \ldots, X_n$, n in number, are linearly independent in $V_n(K)$. Hence A is similar to a diagonal matrix D. The diagonal elements of D are the eigen values $\lambda_1, \ldots, \lambda_n$, by IX.2.1

EXAMPLE 2 : The three eigen values, viz. 1, 2, 3, of the matrix A in example 1(i), page 171, are distinct (real numbers). Hence A is similar to the diagonal matrix D, having 1, 2 and 3 as its diagonal elements.

The eigen vectors, corresponding to 1, 2 and 3, have been found to be $r(1,2,2)$, $r(1,1,1)$ and $r(1,2,1)$ respectively, where r assumes all non-zero real values. Hence, it follows, by IX 2.3, that

$$PAP^{-1}=D=\begin{pmatrix} 1 & & \\ & 2 & \\ & & 3 \end{pmatrix}, \quad \text{where } P=\begin{pmatrix} 1 & 2 & 2 \\ 1 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix}.$$

The condition, laid down in IX.2.5, though sufficient, is *not necessary* for a matrix to be diagonable. This can be seen from the following example.

EXAMPLE 3 : The matrix G given in example 1(iii), page 171, has two distinct eigen values 1 and 5 (of which 1 is 2-fold, and 5 is simple). The eigen vectors, corresponding to the eigen value 1, are given by the non-null vectors, in the linear sub-space, generated by the vectors $(-1, 1, 0)$ and $(0,0,1)$. The eigen vectors, corresponding to the eigen value 5, are given by $r(1, 1, 0)$, where r assumes all non-zero real values. Thus there exist three eigen vectors, $(-1, 1, 0)$, $(0, 0, 1)$ and $(1,1,0)$, which are linearly independent (over the field of real numbers). Hence, by IX.2 4, the matrix G is similar to a diagonal matrix. In fact, it follows by IX.2.8 that

$$PGP^{-1}=\begin{pmatrix}1 & & \\ & 1 & \\ & & 5\end{pmatrix}, \quad \text{where } P=\begin{pmatrix}-1 & 1 & 0\\ 0 & 0 & 1\\ 1 & 1 & 0\end{pmatrix}.$$

The three eigen values of G, viz. 1, 1, 5, are not distinct ; yet G is similar to a diagonal matrix.

A few other necessary and sufficient conditions for a matrix to be diagonable are obtained below.

IX.2.6. *A square matrix A is similar to a diagonal matrix if, and only if, for each eigen value λ of A, the multiplicity of λ is equal to the rank of the characteristic sub-space corresponding to λ (that is, the algebraic and the geometric multiplicities of λ are the same).*

Proof: Let A be a matrix of order n, over a field K.

Let A be similar to a diagonal matrix $D=(\lambda_i)$. Then $PAP^{-1}=D$, where P is a non-singular matrix. The diagonal elements, $\lambda_1, \ldots, \lambda_n$, of D are the eigen values of A. Let λ be an r-fold eigen value of A, so that exactly r of the λ_i are equal to λ. Then $D-\lambda I$ is a diagonal matrix with exactly r diagonal elements equal to 0. Consequently, rank $(D-\lambda I)=n-r$.

Again, since $A-\lambda I=P(D-\lambda I P^{-1}$, and the rank of a matrix remains unaltered on multiplication by any non singular matrix, it follows that rank$(A-\lambda I)=$ rank$(D-\lambda I)=n-r$. Hence the rank of the characteristic sub-space of A corresponding to λ $=$ the nullity of $A-\lambda I=n-(n-r=r$.

Conversely, let the rank of the invariant sub space, corresponding to an eigen value λ, be equal to the multiplicity of λ, for each eigen value λ of A. Let $\mu_1, \ldots, \mu_s$ be the distinct eigen values of A with the respective multiplicities $r_1, \ldots r_s$; then $r_1+\ldots+r_s=n$. Let $W_1, \ldots, W_s$ be the characteristic sub-spaces, corresponding to the eigen values $\mu_1, \ldots, \mu_s$ respectively ; then rank $(W_i)=r_i$, for $i=1,2, \ldots, s$. Let $X_{i1}, X_{i2}, \ldots, X_{ir_i}$ form a basis of W_i, for $i=1,2, \ldots, s$ The vectors X_{ij} are then linearly independent, over K. If possible, let $\Sigma c_{ij} X_{ij}=\theta$, where $c_{ij} \in K$;

i.e.. $\sum_{i=1}^{s}(c_{i1}X_{i1}+\ldots+c_{ir_i}X_{ir_i})=\theta$, or, $Y_1+\ldots+Y_s=\theta$, ... (1)

where $Y_i=c_{i1}X_{i1}+\ldots+c_{ir_i}X_{ir_i}$, for $i=1,2,\ldots,s$.

Then $Y_i \varepsilon W_i$, for $i=1,2,\ldots,s$. As the eigen values $\mu_1,\ldots,\mu_s$ are distinct, and $W_1,\ldots,W_s$ are the corresponding characteristic sub-spaces, it follows, by VIII.5.4, that the vectors $Y_1,\ldots,Y_s$ are linearly independent, if they are non-null Hence, it follows from (1) that each of $Y_1,\ldots,Y_s$ is necesarily the null vector, so that

$$c_{i1}X_{i1}+\ldots+c_{ir_i}X_{ir_i}=\theta, \text{ for } i=1,2,\ldots s. \quad \ldots (2)$$

Again, since the vectors $X_{i1},\ldots,X_{ir_i}$ form a basis of W_i, they are linearly independent over K. Hence, it follows from (2) that $c_{ij}=0$, for $j=1,\ldots r_i$ and $i=1,2,\ldots,s$.

Thus the vectors X_{ij} are linearly independent over K. And since they are $r_1+\ldots+r_s=n$ in number, they form a basis of $V_n(K)$. Hence, by IX.2.4, A is similar to a diagonal matrix.

IX.2.7. *A square matrix A of order n, over a field K, is similar to a diagonal matrix if, and only if, the elementary divisors of the characteristic matrix $A-xI$ in $K[x]$ be all of the first degree.*

Proof: Let A be similar to a diagonal matrix $D=(d_i)$. The elementary divisors of $D-xI$ are given by $d_1-x, d_2-x,\ldots,d_n-x$. Hence, by Corollary 2 of IX.1.3, the elementary divisors of $A-xI$ are the linear polynomials $d_1-x,\ldots,d_n-x \;\varepsilon\; K[x]$.

Conversely, let the elementary divisors of $A-xI$ be the linear polynomials $d_1-x,\ldots,d_n-x \;\varepsilon\; K[x]$. Then for the diagonal matrix $D=(d_i)$, the characteristic matrix $D-xI$ has also the same set of elementary divisors. And, also $D-xI$ and $A-xI$ have the same rank n. Hence, by Corollary 2 of IX.1.3, A is similar to D, which is a diagonal matrix.

It has been shown in page 129 how the invariant factors of a matrix, say $A-xI$, can be obtained from its elementary divisors In fact, the last invariant factor is obtained as the product of the highest powers of the distinct irreducible elements (i.e, polynomials in the present case) among the elementary divisors. Hence, if the elementery divisors of $A-xI$ are linear polynomials in $K[x]$, the last invariant factor of $A-xI$ is a product of distinct linear factors ; and conversely. Also the last invariant factor of $A-xI$ is the minimal polynomial of A Thus one obtains the following from IX.2.7 :

IX.2.8. *A square matrix A over a field K, is similar to a diagonal matrix if, and only if, the minimal polynomial of A is a product of distinct linear factors in $K[x]$.*

Exercises

1. Find the algebraic and the geometric multiplicities of the eigen values of the following matrices. Hence find which of these are diagonable :

(i) $\begin{pmatrix} 1 & -1 & 0 \\ 1 & 2 & 1 \\ -2 & 1 & -1 \end{pmatrix}$, (ii $\begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 1 & 1 \end{pmatrix}$, (iii) $\begin{pmatrix} 1 & 0 & -1 \\ 2 & 2 & 2 \\ 2 & 1 & 2 \end{pmatrix}$,

(iv) $\begin{pmatrix} 3 & 2 & 1 & 2 \\ 2 & 3 & 1 & 2 \\ 2 & 2 & 2 & 2 \\ -4 & -1 & -1 & -1 \end{pmatrix}$, (v) $\begin{pmatrix} -1 & 1 & -1 & -1 \\ -1 & -2 & 1 & -1 \\ -6 & -3 & 0 & -5 \\ 3 & 0 & 1 & 3 \end{pmatrix}$.

2 Find which of the matrices in exercise 1 are diagonable by applying the propositions IX 2.4, IX.2.7 and IX.2.8.

3. Block-diagonal matrices.

Let A be a square matrix of order n, over a field K. Then A is similar to a diagonal matrix, only when some particular conditions are satisfied. Thus A need not, in general, be similar to a diagonal matrix. To obtain normal forms of matrices, under similarity, one has therefore to consider a class of matrices, wider than that of the diagonal matrices. The block diagonal matrices are found to be suitable for the purpose.

A matrix of the form,

$$A = \begin{pmatrix} A_1 & O & \dots & O \\ O & A_2 & \dots & O \\ \dots & \dots & \dots & \dots \\ O & O & \dots & A_s \end{pmatrix},$$

where $A_1, A_2, \dots, A_s$ are square matrices (not necessarily of the same order) and each O is a rectangular matrix of suitable size, having all its elements null, is called a *block-diagonal matrix*; $A_1, A_2 \dots, A_s$ are called the *blocks* of A. Then the matrix A is usually expressed as

$$A = \text{diag}\ (A_1, A_2, \dots, A_s). \qquad \dots \ (1)$$

A matrix A, which is expressible as a block-diagonal matrix (1), and for which $s > 1$, is said to be *decomposable* into the submatrices $A_1, \dots, A_s$; otherwise, A is said to be an *indecomposable* matrix.

Now, one obtains readily the following, by the laws of block-multiplication of matrices (vide pages 59—60):

IX.3.1. *Let* $A = diag(A_1, \dots, A_s)$ *and* $B = diag(B_1, \dots B_s)$, *where* A_i *and* B_i *are of the same order, for* $i = 1, \dots, s$, *then*

(a) $A + B = diag\ (A_1 + B_1, \dots, A_s + B_s)$,

(b) $AB = diag\ (A_1B_1, \dots, A_sB_s)$, *and*

(c) $gA = diag\ (gA_1, \dots, gA_s)$, *where* g *is any element.*

(d) *If* A *be non-singular, then*

$$A^{-1} = diag\ (A_1^{-1}, \dots, A_s^{-1}).$$

As an immediate consequence of the above, one obtains:

IX.3.2. (i) *If* $A = diag\ (A_1, \dots, A_s)$, *then*

(a) $f(A) = diag\ (f(A_1), \dots, f(A_s))$, *where* $f(x) \varepsilon K[x]$;

(b) $A - xI = diag\ (A_1 - xI, \dots, A_s - xI)$.

(ii) *If the matrices* A_i *be similar to the matrices* B_i, *for* $i = 1, \dots, s$, *then* $A = diag(A_1, \dots, A_s)$ *is similar to* $B = diag\ (B_1, \dots, B_s)$.

(iii) *If* A_i *be equivalent to* B_i, *for* $i, \dots, s$, *over* $K[x]$, *where* K *is a field, then* $A = diag\ (A_1, \dots, A_s)$ *is equivalent to* $B = diag\ (B_1, \dots, B_s)$, *over* $K[x]$.

Proof: (i) The result follows immediately, by IX.3.1.

(ii) Let $B_i = P_iA_iP_i^{-1}$, for $i=1,\ldots,s$, then (by IX.3.1)

$$B = \text{diag}\,(B_1,\ldots,B_s) = \text{diag}\,(P_1A_1P_1^{-1},\ldots,P_sA_sP_s^{-1})$$
$$= PAP^{-1}, \quad \text{where } P = \text{diag}\,(P_1,\ldots,P_s).$$

(iii) Let $B_i = P_iA_iQ_i$, where P_i, Q_i are products of unimodular elementary matrices over $K[x]$, for $i=1,\ldots,s$. Then

$$B = \text{diag}\,(B_1,\ldots,B_s) = \text{diag}\,(P_1A_1Q_1,\ldots,P_sA_sQ_s) = PAQ.$$

where $P = \text{diag}\,(P_1,\ldots,P_s)$ and $Q = \text{diag}\,(Q_1,\ldots,Q_s)$ are also products of unimodular elementary matrices over $K[x]$.

IX.3.3. *The characteristic polynomial of a block-diagonal matrix is the product of the characteristic polynomials of its diagonal blocks.*

Proof: Let $A = \text{diag}\,(A_1,\ldots,A_s)$ then (by IX.3.2-i(b)), $A - xI = \text{diag}\,(A_1 - xI,\ldots,A_s - xI)$. Hence

$$\psi_A(x) = \det(A - xI) = \det(A_1 - xI)\ldots\det(A_s - xI)$$
$$= \psi_{A_1}(x)\ldots\psi_{A_s}(x).$$

IX.3.4. *The minimal polynomial of a block-diagonal matrix is the least common multiple of the minimal polynomials of the diagonal blocks.*

Proof: Let $A = \text{diag}\,(A_1,\ldots,A_s)$ be a block-diagonal matrix, over a field K. Let $m(x)$, $m_1(x),\ldots,m_s(x)$ denote the minimal polynomials of the matrices A, $A_1,\ldots,A_s$ respectively; and let $f(x)$ be the least common multiple of $m_1(x),\ldots,m_s(x)$ in $K[x]$. By IX.3.2-*i*(*a*),

$$m(A) = \text{diag}\,(m(A_1),\ldots,m(A_s)). \qquad \ldots (2)$$

As $m(A) = O$ (the null matrix), it follows from (2) that $m(A_1) = O,\ldots,m(A_s) = O$. Hence, $m(x)$ is divisible by each of the polynomials $m_1(x),\ldots,m_s(x)$, by VIII.3.1. Thus $m(x)$ is a common multiple of the polynomials $m_1(x),\ldots,m_s(x)$, and is therefore divisible by $f(x)$.

On the otherhand, $f(A_1) = \ldots, = f(A_s) = O$ (by VIII.3.1). Consequently, $f(A) = \text{diag}\,(f(A_1),\ldots,f(A_s)) = O$; and $f(x)$ is therefore divisible by $m(x)$ (by VIII.3.1).

Thus $f(x)$ and $m(x)$ are mutually divisible in $K[x]$. Hence $m(x)$ is the least common multiple of $m_1(x),\ldots,m_s(x)$.

IX 3.5. *The set of elementary divisors of a block-diagonal matrix, over $K[x]$, where K is a field, is the union of sets of elementary divisors of its diagonal blocks.*

Proof : Let $A=\text{diag}(A_1, \ldots, A_s)$ be a block-diagonal matrix, over $K[x]$, where K is a field. Elementary row and column operations on the separate blocks $A_1, \ldots, A_s$ may be considered as elementary operations on the whole matrix. And such operations applied on one block does not affect any other block. The blocks $A_1, \ldots, A_s$ are reduced to their respective Smith's normal forms $S_1, \ldots, S_s$, by suitable elementary operations. The matrix A is, thereby, reduced to the matrix $D=\text{diag}(S_1, \ldots, S_s)$. As each of $S_1, \ldots, S_s$ is a diagonal matrix, D is also a diagonal matrix. The diagonal elements of S_i are the invariant factors of A_i ; and the prime-power factors of the diagonal elements are the elementary divisors of A_i, for $i=1, \ldots, s$. Thus the prime-power factors of the diagonal elements of D constitute the union of the sets of elementary divisors of the diagonal blocks $A_1, \ldots, A_s$

Again, the prime-power factors of the diagonal elements of the diagonal matrix D are the elementary divisors of D (by exercise 1, page 131). And, as A is equivalent to D, and both have the same rank, those are also the elementary divisors of A.

Let $A=\text{diag}(A_1, \ldots, A_s)$ be a block-diagonal matrix. The positions of two blocks may be interchanged by interchanging some rows and the corresponding columns. To interchange the ith and the jth rows of a matrix A, one has to pre-multiply A by the elementary matrix I_{ij} ; and to interchange the ith and the jth columns, one has to post-multiply by I_{ij}. But, since $I_{ij}= I_{ij}^{-1}$, it follows that the matrix, obtained by interchanging some rows and the corresponding columns, is similar to the given matrix. Thus one obtains the following :

IX.3.6. *If A and B be two block-diagonal matrices, over a field K, both having the same diagonal blocks, but in different arrangements, then A and B are similar.*

Again, if the rows and columns of a matrix, over a Euclidean domain be interchanged, the resulting matrix is equivalent to the given matrix. Hence, it follows that :

IX.3.7. *If A and B be two block-diagonal matrices, over $K[x]$, where K is a field, both A and B having the same diagonal blocks, but in different arrangements, then A is equivalent to B, over $K[x]$.*

EXAMPLE 1: *A square matrix A, over a field K, is similar to a (lower) triangular matrix whose diagonal elements are the eigen values of A (provided that all the eigen values lie in K).*

Soln : Let A be of order n, and let $\lambda_1,\ldots,\lambda_n$ be the eigen values of A (lying in K). Let X_1 be an eigen vector of A corresponding to the eigen value λ_1. Let P_1 be any non-singular matrix having X_1 as its first row-vector Then the first row-vector of P_1A

$$=X_1A=\lambda_1X_1=\text{the first row-vector of } P_1.\lambda_1 I.$$

Hence, the first row-vector of $P_1AP_1^{-1}$

$$\begin{aligned}&=(\text{the first row-vector of } P_1A).\,P_1^{-1}\\&=(\text{the first row-vector of } P_1.\,\lambda_1 I).\,P_1^{-1}\\&=\text{the first row-vector of } P_1.\,\lambda_1 I.\,P_1^{-1}\\&=\text{the first row-vector of } \lambda_1 I\\&=(\lambda_1, 0,\ldots,0).\end{aligned}$$

Thus $$P_1AP_1^{-1}=\begin{pmatrix}\lambda_1 & O\\ B_1 & A_1\end{pmatrix},\quad \text{where } A_1 \text{ is of order } n-1.$$

Again, since $\psi_A(x)=\det(A-xI)=\det(P_1AP_1^{-1}-xI)=(\lambda_1-x).\det(A_1-xI)$, it follows that $\lambda_2,\ldots,\lambda_n$ are the eigen values of A_1.

If $n=2$, then A_1 is the 1×1 matrix (λ_2) and A is similar to the triangular matrix $\begin{pmatrix}\lambda_1 & O\\ b_1 & \lambda_2\end{pmatrix}$.

If $n>2$, then one finds, as above, a non-singular matrix P_2 (of order $n-1$), such that

$$P_2A_1P_2^{-1}=\begin{pmatrix}\lambda_2 & O\\ B_2 & A_2\end{pmatrix},\quad \text{where } A_2 \text{ is of order } n-2, \text{ and such}$$

that $\lambda_3,\ldots,\lambda_n$ are the eigen values of A_2.

The above process is applied $(n-1)$–times to obtain the final result. The matrix

$$P=\begin{pmatrix}I_{n-2} & O\\ O & P_{n-1}\end{pmatrix}\cdots\begin{pmatrix}I_2 & O\\ O & P_3\end{pmatrix}\begin{pmatrix}I_1 & O\\ O & P_2\end{pmatrix}P_1,$$

where I_r is the identity matrix of order r, is then a non-singular matrix of order n, such that PAP^{-1} is a lower triangular matrix, having $\lambda_1,\ldots,\lambda_n$ as its diagonal elements.

Exercises

1. If the characteristic polynomial of a square matrix A, over a field K, be the product of two polynomials $f(x)$ and $g(x)$, which are relatively prime in $K[x]$, then A is similar to a block diagonal matrix, $B=\text{diag}(B_1, B_2)$, where $f(x)$ and $g(x)$ are the characteristic polynomials of the blocks B_1 and B_2.

2. If the minimal polynomial of a square matrix A, over a field K, be the product of two polynomials $f(x)$ and $g(x)$, which are relatively prime in $K[x]$, then A is similar to a block diagonal matrix, $C=\text{diag}(C_1,C_2)$, where $f(x)$ and $g(x)$ are the minimal polynomials of the blocks C_1 and C_2.

3. A square matrix A, over a field K, is indecomposable over K if, and only if, the characteristic and the minimal polynomials of A are the same and have the form $[f(x)]^r$, where $f(x)$ is irreducible in $K[x]$.

4. Normal forms.

Different types of normal forms for square matrices, over a field, under similarity, may now be obtained.

Jordan normal form (or classical canonical form). Matrices of a particular type, known as Jordan matrices, can be utilised as normal forms, which are canonical, for matrices under similarity.

A decomposable matrix, $J=\text{diag}\ (J_1(\lambda_1),\ldots, J_s(\lambda_s))$, for which the diagonal blocks $J_i(\lambda_i)$ are Jordan blocks (vide page 160) whose orders may be different and $\lambda_1,\ldots,\lambda_s$ may not be distinct, is called a *Jordan matrix.*

IX.4.1. *A square matrix A, over a field K, such that the eigen values of A lie in K, is similar to a Jordan matrix, which is determined uniquely, except for rearrangents of its diagonal blocks.*

Proof: Let A be a square matrix of order n, over K. Since the eigen values of A lie in K, the characteristic polynomial of A decomposes into linear factors in $K[x]$,

$$\psi_A(x)=(\lambda_1-x)^{n_1}(\lambda_2-x)^{n_2}\ldots(\lambda_m-x)^{n_m},\quad n_1+\ldots+n_m=n,$$

where $\lambda_1,\ldots,\lambda_m$ are the distinct eigen values of A. Again $\psi_A(x)$ is the product of the invariant factors, and hence also of the elementary divisors, of the characteristic matrix $A-xI$. As the prime factors of $\psi_A(x)$ in $K[x]$ are the linear polynomials λ_i-x, it follows that the elementary divisors of $A-xI$ are of the form $(\lambda_i-x)^{r_i}$. Let $(\lambda_1-x)^{r_1},\ldots,(\lambda_s-x)^{r_s}$ be the complete set of elementary divisors of $A-xI$, where $\lambda_1,\ldots,\lambda_s$ may not be distinct.

Let $J_i(\lambda_i)$ denote the Jordan block of order r_i, for $i=1,\ldots,s$. Then, by Corollary 3, page 160, $(\lambda_i-x)^{r_i}$ is both the characteristic plynomial and the minimal polynomial of $J_i(\lambda_i)$. As the minimal polynomial of $J_i(\lambda_i)$ is the last invariant factor of $J_i(\lambda_i)-xI$, and the characteristic polynomial of $J_i(\lambda_i)$ is the product of all the invariant factors of $J_i(\lambda_i)-xI$, it follows that all the invariant factors of $J_i(\lambda_i)-xI$, excepting the last one, is 1; consequently $(\lambda_i-x)^{r_i}$ is the only elementary divisor of $J_i(\lambda_i)-xI$.

Let J denote the Jordan matrix, $J=\text{diag}(J_1(\lambda_1),\ldots,J_s(\lambda_s))$. Then by IX.3.5, the set of the elementary divisors of $J-xI$, is given by the union of the elementary divisors of $J_1(\lambda_1)-xI,\ldots,J_s(\lambda_s)-xI$. Thus the elementary divisors of $J-xI$ are given by $(\lambda_1-x)^{r_1},\ldots,(\lambda_s-x)^{r_s}$.

Again, since the product of the elementary divisors of $A-xI$ is equal to the characteristic polynomial $\psi_A(x)$, it follows that $r_1+\ldots+r_s=n_1+\ldots+n_m=n$; consequently $J-xI$ has the order n. Also, as $A-xI$ and $J-xI$ are both non-singular, they have the same rank.

Thus $A-xI$ and $J-xI$ are two square matrices of order n, possessing the same rank and the same set of elementary divisors. Hence, by Corollary 2 of IX.1.3, the matrix A is similar to the Jordan matrix J.

The set of elementary divisors of $A-xI$ consists of the elementary divisors of $J_i(\lambda_i)-xI$, for $i=1,\ldots,s$, there being one elementary divisor corresponding to each block $J_i(\lambda_i)$. Also two block-diagonal matrices, which differ only in the arrangement of diagonal blocks, are similar.

Hence the Jordan matrix J is determined uniquely, except for rearrangements of the diagonal blocks, by the given matrix A.

The Jordan matrix J, similar to A, is called the *Jordan normal form* (or the *classical canonical form*) of the matrix A.

If the ground field K be, in particular, an algebraically closed field (e.g. the field of complex numbers), then all the eigen values of any matrix, over K, lie in K. Thus one obtains the following :

IX.4.2. *Any square matrix, over an algebraically closed field, is similar to a Jordan matrix, which is determined uniquely, except for rerrangements of its diagonal blocks.*

To find the Jordan's normal form of a given square matrix A, over a field K, which contains all the eigen values of A, the elementary divisors of $A-xI$ are found. Let $(\lambda_i-x)^{r_{it}}$, for $t=1,\ldots,k_i$, where $i=1,\ldots,m$, be the elementary divisors of the characteristic matrix $A-xI$. Then the multiplicities of the eigen values λ_i, as roots of these elementary divisors i.e., as the roots of the invariant factors of $A-xI$, are collected together separately in first brackets, for each one of the eigen values, as follows :

$$\lambda_1 : (r_{11},\ldots,r_{1k_1}),\ \lambda_2 : (r_{21},\ldots,r_{2k_2}),\ldots,\ \lambda_m : (r_{m1},\ldots,r_{mk_m}).$$

Then the symbol

$$[(r_{11},\ldots,r_{1k_1}),\ (r_{21},\ldots,r_{2k_2}),\ldots,\ (r_{m1},\ldots,r_{mk_m})]$$

is called *the Segre characteristic of the matrix A*. It is customary to arrange $r_{i1},\ldots,r_{ik_i}$ in non-increasing order of magnitude. The Jordan matrix, similar to A, has then one Jordan block of order r_{it}, for each r_{it} (for $t=1,\ldots,k_i$ and $i=1,\ldots,m$), with the corresponding eigen value λ_i.

EXAMPLE 1 ; Let A be a real square matrix of order 11 ; and let the invariant factors of $A-xI$ be

$$e_1(x)=\ldots=e_8(x)=1,\quad e_9(x)=2-x,$$
$$e_{10}(x)=(2-x)(6+x)^2,\ \text{and}\ e_{11}(x)=(2-x)^2(6+x)^2(5-x)^3.$$

Then the elementary divisors of $A-xI$ are given by

$$(2-x)^2,\ (2-x),\ (2-x),\ (6+x)^2,\ (6+x)^2, (5-x)^3.$$

Hence the Segre characteristic of the matrix A is given by

$$[(2,1,1),(2,2),(3)],$$

having 2, −6 and 5 as the corresponding eigen values. The Jordan normal form is then given by

$$J=\operatorname{diag}(J_1,\ldots,J_6),$$

where $J_2=J_3=(2)$,

$$J_1=\begin{pmatrix}2&1\\0&2\end{pmatrix},\quad J_4=J_5=\begin{pmatrix}-6&1\\&-6\end{pmatrix},\quad\text{and } J_6=\begin{pmatrix}5&1&0\\0&5&1\\0&0&5\end{pmatrix}.$$

EXAMPLE 2 : To find the Jordan normal forms of the following matrices, over the field of real numbers :

$$A=\begin{pmatrix}4&-1&1\\4&0&2\\2&-1&3\end{pmatrix},\quad B=\begin{pmatrix}-6&6&1\\-1&1&1\\-8&5&3\end{pmatrix}.$$

Soln : The characteristic matrices of A and B are given by

$$A-xI=\begin{pmatrix}4-x&-1&1\\4&-x&2\\2&-1&3-x\end{pmatrix},\quad B-xI=\begin{pmatrix}-6-x&6&1\\-1&1-x&1\\-8&5&3-x\end{pmatrix}.$$

The characteristic polynomials of A and B are

$$\psi A(x)=-x^3+7x^2-16x+12=-(x-2)^2(x-3),$$

and $$\psi B(x)=-x^3-2x^2+12x-15=-(x+5)(x^2-3x+3).$$

As the roots of $\psi B(x)$, i.e., the eigen values of B are not all real, *there does not exist any Jordan normal form of the matrix B, over the real field.*

The determinant divisors of $A-xI$ are found to be

$$d_1(x)=1,\quad d_2(x)=x-2,\text{ and } d_3(x)=(x-2)^2(x-3).$$

Hence the invariant factors of $A-xI$ are given by

$$e_1(x)=d_1(x)=1,\quad e_2(x)=d_2(x)/d_1(x)=x-2,$$

and $$e_3(x)=d_3(x)/d_2(x)=(x-2)(x-3).$$

The multiplicities of the eigen values 2 and 3, as roots of the invariant factors, are now tabulated as follows :

	$e_3(x)$	$e_2(x)$	$e_1(x)$
2	1	1	0
3	1	0	0

The Segre characteristic of the matrix is thus found to be

$$[(1,1),(1)].$$

The Jordan normal form of the matrix A is correspondingly given by :

$$\begin{pmatrix}2&&\\&2&\\&&3\end{pmatrix}.$$

Note : The matrix A is similar to a diagonal matrix, although its eigen values are not distinct.

EXAMPLE 3 : Let A be a square matrix of order 3, over a field K, such that the eigen values of A lie in K. To find the different forms of Jordan matrices similar to A.

Soln : Let e_1, e_2, e_3 be the invariant factors of the characteristic matrix $A-xI$. Then (a) e_1 is a divisor of e_2, and e_2 is a divisor of e_3, and (b) their product $e_1e_2e_3=\psi A(x)$, the characteristic polynomial of the matrix A. The properties (a) and (b) enable one to obtain the possible different cases that may arise.

(I) Let there exist three distinct eigen values λ_1, λ_2, λ_3. Then

$$\psi A(x)=(\lambda_1-x)(\lambda_2-x)(\lambda_3-x).$$

Hence, by (a) and (b),

$e_1=1$, $e_2=1$, $e_3=(\lambda_1-x)(\lambda_2-x)(\lambda_3-x)$,

so that the Segre characteristic is $[(1), (1), (1)]$, and the Jordan normal form is given by the matrix T_1.

$$T_1=\begin{pmatrix}\lambda_1 & 0 & 0\\ 0 & \lambda_2 & 0\\ 0 & 0 & \lambda_3\end{pmatrix}.$$

(II) Let there be two distinct eigen values λ_1 and λ_2, of which λ_1 is 2-fold. Then

$$\psi A(x)=(\lambda_1-x)^2(\lambda_2-x).$$

Two different cases arise :

(i) $e_1=1$, $e_2=\lambda_1-x$, $e_3=(\lambda_1-x)(\lambda_2-x)$,

so that the Segre characteristic is $[(1,1), (1)]$, and the Jordan normal form is given by the matrix T_2.

$$T_2=\begin{pmatrix}\lambda_1 & 0 & 0\\ 0 & \lambda_1 & 0\\ 0 & 0 & \lambda_2\end{pmatrix}.$$

(ii) $e_1=1$, $e_2=1$. $e_3=(\lambda_1-x)^2(\lambda_2-x)$,

so that the Segre characteristic is $[(2), (1)]$, and the Jordan normal form is given by the matrix T_3.

$$T_3=\begin{pmatrix}\lambda_1 & 1 & 0\\ 0 & \lambda_1 & 0\\ 0 & 0 & \lambda_2\end{pmatrix}.$$

(III) Let there be only one distinct eigen value λ, which is of multiplicity 3. Then

$$\psi A(x)=(\lambda-x)^3.$$

Three different cases arise :

(i) $e_1=\lambda-x$, $e_2=\lambda-x$, $e_3=\lambda-x$,

so that the Segre characteristic is $[(1, 1, 1)]$, and the Jordan normal form is given by the matrix T_4.

$$T_4=\begin{pmatrix}\lambda & 0 & 0\\ 0 & \lambda & 0\\ 0 & 0 & \lambda\end{pmatrix}.$$

(ii) $e_1=1, e_2=\lambda-x, e_3=(\lambda-x)^2$

so that the Segre characteristic is $[(2, 1)]$, and the Jordan normal form is given by the matrix T_5.

$$T_5=\begin{pmatrix}\lambda & 1 & 0\\ 0 & \lambda & 0\\ 0 & 0 & \lambda\end{pmatrix},$$

(iii) $e_1=1, e_2=1, e_3=(\lambda-x)^3$,

so that the Segre characteristic is $[(3)]$, and the Jordan normal form is given by the matrix T_6.

$$T_6=\begin{pmatrix}\lambda & 1 & 0\\ 0 & \lambda & 1\\ 0 & 0 & \lambda\end{pmatrix}$$

Note : It may be readily verified, by finding the rank of $T_i-\lambda I$, for $i=1,\ldots,6$, that, over the field of real numbers, the matrices $T_1,\ldots,T_6$ are the Jordan normal forms of the matrices of the collineations described in cases (I), (II) and (III) in pages 176-177. But there does *not* exist any Jordan normal form corresponding to case (IV).

A square matrix A, over an arbitrary ground field K, may *not* be similar to any Jordan matrix ; that is, *there may not exist any Jordan normal form* for the matrix A. There exist other types of normal forms to which any matrix, over an arbitrary ground field, can always be reduced. Some such normal forms will now be obtained.

Natural normal form (or rational canonical form). Let A be a square matrix of order n, over a field K. Let

$$e_1(x)=\ldots=e_m(x)=1,\ e_{m+1}(x)=g_1(x),\ldots,e_n(x)=g_s(x),$$

(where $s=n-m$) be the invariant factors of the characteristic matrix $A-xI$; and let the leading coefficients in $g_1(x),\ldots,g_s(x)$ be 1. Then $g_i(x)$ is a divisor of $g_{i+1}(x)$ in $K[x]$, for $i=1,\ldots,s-1$.

Let, now, B be the block-diagonal matrix,

$$B=\text{diag}(B_1,\ldots,B_s),$$

where B_i is the companion matrix of the polynomial $g_i(x)$, for $i=1,\ldots,s$. Then the order of $B=$ sum of the degrees of $g_1(x),\ldots g_s(x)=$ degree of $\psi_A(x)=n$ [since $\psi_A(x)=g_1(x)\ldots g_s(x)$].

Again, since the characteristic polynomial, $(-1)^{r_i}.g_i(x)$ of B_i is also the minimal polynomial of B_i (r_i being the degree of $g_i(x)$), it follows that the invariant factors of the characteristic matrix B_i-xI are given by $1,\ldots,1,\ g_i(x)$ (the invariant factors being determined uniquely except for their signs), for $i=1,\ldots,s$. Also, $B-xI=\text{diag}\,(B_1-xI,\ldots,B_s-xI)$. Elementary row and column operations are applied to reduce the blocks $B_1-xI,\ldots,B_s-xI$ to their respective Smith's normal forms over $K[x]$ (any elementary operation on a particular block B_i-xI may be considered as an elementary operation on the whole matrix $B-xI$, and it does not affect any other block). The matrix $B-xI$ is, thereby, reduced to a diagonal matrix, whose diagonal elements are arranged as

$$1,\ldots,1, g_1(x), 1,\ldots,1, g_2(x), \ldots, 1,\ldots,1, g_s(x)$$

By suitable interchanges of rows and columns, the matrix is is further reduced to the diagonal matrix D, whose diagonal elements are arranged as

$$1,\ldots,1, g_1(x),\ldots,g_s(x). \qquad \ldots(1)$$

Since $g_i(x)$ is a divisor of $g_{i+1}(x)$, for $i=1,\ldots,s-1$, it follows that D is the Smith's normal form of the matrix $B-xI$, over $K[x]$. Thus the invariant factors of $B-xI$ are given by (1).

Since A and B are square matrices of the same order n, over K, and $A-xI$ and $B-xI$ have the same set of invariant factors (1), over $K[x]$, it follows (by Corollary 1 of IX.1.3) that A is similar to the matrix B. Thus one obtains the following :

IX.4.3. *Any square matrix A, over a field K, is similar to a block-diagonal matrix, $B=\text{diag}\,(B_1,\ldots,B_s)$, where $B_1,\ldots,B_s$ are the companion matrices of the invariant factors, different from 1, of the characteristic matrix $A-xI$ (over $K[x]$).*

The matrix B, as defined in IX.4,3, is determined uniquely, except for rearrangements of diagonal blocks, and it is called *the natural normal form* (or the *rational canonical form*) of A.

Quasi-natural normal form. The natural normal form always exists for any square matrix, over any field whatsoever. But its diagonal blocks are usually found to be decomposable; and then it is possible to reduce it further, and obtain the quasi-natural normal form, whose diagonal blocks are indecomposable and smaller in size. In fact, it follows readily from IX.3.5 and Corollary 2, page 160, that :

IX.4.4. *Any square matrix A, over a field K is similar to a block-diagonal matrix, $C=\text{diag}\,(C_1,\ldots,C_m)$, where $C_1,\ldots,C_m$*

are the companion matrices of the elementary divisors (with their leading coefficients taken equal to 1) of the characteristic matrix $A-xI$ *over* $K[x]$.

The matrix C, as defined in IX.4.4, is determined uniquely, except for rearrangements of its diagonal blocks ; it is called the *quasi-natural normal form* of the given matrix A. The blocks C_i are indecomposable over K ; otherwise, the characteristic polynomial of the block C_i would yield more than one elementary divisor of C_i-xI, over $K[x]$.

EXAMPLE 4 : Let A be a square matrix of order 10, over the field of rational numbers R, such that the invariant factors of $A-xI$, over R$[x]$, are given by

$$e_1(x)=\ldots=e_7(x)=1,\quad e_8(x)=x-2,\quad e_9(x)=(x-2)(x^2+x+1)=x^3-x^2-x-2,$$

$$e_{10}(x)=(x-2)^2(x^2+x+1)(x^2+2x+4)=x^6-x^5-x^4-10x^3+8x^2+8x+16.$$

Let B_1, B_2 and B_3 denote the companion matrices of $e_8(x)$, $e_9(x)$ and $e_{10}(x)$ respectively. Then

$$B_1=(2),$$

$$B_2=\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & 1 & 1 \end{pmatrix},\qquad B_3=\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -16 & -8 & -8 & 10 & 1 & 1 \end{pmatrix},$$

and the natural normal form of the matrix A is given by the block-diagonal matrix, $B=\text{diag}\,(B_1, B_2, B_3)$.

The elementary divisors of $A-xI$, over R$[x]$, are

$$x-2,\ x-2,\ (x-2)^2,\ x^2+x+1,\ x^2+x+1,\ \text{and}\ x^2+2x+4.$$

and the corresponding companion matrices are

$$C_1=C_2=(2),\quad C_3=\begin{pmatrix} 0 & 1 \\ -4 & 4 \end{pmatrix},\quad C_4=C_5=\begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix},\ \text{and}\ C_6=\begin{pmatrix} 0 & 1 \\ -4 & -2 \end{pmatrix}.$$

The quasi-natural normal form of the matrix A is given by the block-diagonal matrix, $C=\text{diag}\,(C_1, C_2, C_3, C_4, C_5, C_6)$.

Generalised Jordan normal form (or Jacobson canonical form). Let K be a *perfect* field, and let $f(x)$ be a monic polynomial (of degree m), irreducible in $K[x]$. Also, let M be a matrix, over K, having $f(x)$ as its characteristic polynomial. Then the matrix,

$$G=\begin{pmatrix} M & I & O & \ldots & O & O \\ O & M & I & \ldots & O & O \\ \ldots & \ldots & & \ldots & \ldots & \\ O & O & O & \ldots & M & I \\ O & O & O & \ldots & O & M \end{pmatrix}, \qquad \ldots (2)$$

with r rows, is called *the generalised Jordan block* corresponding to $[f(x)]^r$, I being the identity matrix (of proper order). If M be taken, in particular, to be the companion matrix of $f(x)$, then G is also called *the hyper-companion matrix* corresponding to $[f(x)]^r$.

Since the characteristic polynomial $(-1)^m f(x)$ of the matrix M is irreducible in $K[x]$, it follows that the invariant factors of $M-xI$ in $K[x]$ are given by

$$e_1(x)=\dots=e_{m-1}(x)=1, \text{ and } e_m(x)=f(x)$$

(the invariant factors may be taken with arbitrary sign); hence $M-xI$ has only a single elementary divisor $f(x)$ in $K[x]$.

Now, the characteristic matrix of G is

$$G-xI=\begin{bmatrix} H & I & O & \dots & O & O \\ O & H & I & \dots & O & O \\ \dots & \dots & \dots & \dots & \dots & \dots \\ O & O & O & \dots & H & I \\ O & O & O & \dots & O & H \end{bmatrix}, \qquad \dots (3)$$

where $H=M-xI$. The matrix (3) may be reduced to the block-diagonal matrix $C=\operatorname{diag}(I,\dots I, [(-1)^{r-1}H^r])$, by the following elementary row and column operations :

(i) adding $(-H)$-times 2nd column to the first column, then

(ii) adding (H^2) times 3rd column to the first column, and $(-H)$-times 3rd column to the 2nd column, then

(iii) adding $(-H^3)$-times 4th column to the first column, (H^2)-times 4th column to the 2nd column, and $(-H)$-times 4th column to the 3rd column, and

so on upto column operations with the rth column ; and

(α) then adding $(-H)$-times the $(r-1)$-th row, (H^2)-times the $(r-2)$-th row,....,$(1)^{r-2}H^{r-1}$-times the first row to the rth row ; and

finally by interchanging the rows.

By IX.3.5, the elementary divisors of C in $K[x]$ are the same as those of H^r. And, since $G-xI$ is equivalent to C, over $K[x]$, the elementary divisors of $G-xI$, in $K[x]$, are given by those of H^r.

Let Γ be the splitting field of $f(x)$, over K (or Γ may be taken to be the algebraic closure of K). As K is a perfect field, and $f(x)$ is irreducible in $K[x]$, it follows that $f(x)$ is separable. Hence the roots $\lambda_1, \ldots, \lambda_m$ of $f(x)$ in Γ are distinct. Now, since the eigen values of M are distinct, M is similar to the diagonal matrix with $\lambda_1, \ldots, \lambda_m$ as the diagonal elements, *over* Γ. Hence there exists a non-singular matrix P, over Γ, such that

$$PMP^{-1} = \begin{pmatrix} \lambda_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \lambda_m \end{pmatrix}, \quad \text{then } PMP^{-1} - xI = \begin{pmatrix} \lambda_1 - x & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \lambda_m - x \end{pmatrix},$$

so that

$$(PMP^{-1} - xI)^r = \begin{pmatrix} (\lambda_1 - x)^r & & & \\ & \cdot & & \\ & & \cdot & \\ & & & (\lambda_m - x)^r \end{pmatrix}.$$

Also, $PH^rP^{-1} = (PHP^{-1})^r = [P(M - xI)P^{-1}]^r = (PMP^{-1} - xI)^r$. Thus H^r is equivalent to the matrix $(PMP^{-1} - xI)^r$, over $\Gamma[x]$ (since the elements of P belong to Γ, and the non-null elements are units in $\Gamma[x]$). Hence, the elementary divisors of H^r, being the same as those of $(PMP^{-1} - xI)^r$, are given by $(x - \lambda_1)^r, \ldots, (x - \lambda_m)^r$. And, since these are relatively prime, the invariant factors of H^r, over $\Gamma[x]$, are given by $1, \ldots 1, [(x - \lambda_1)^r \ldots (x - \lambda_m)^r] = [(f)]^r$. But $f(x)$ is irreducible in $K[x]$; hence H^r has $[f(x)]^r$ as its only elementary divisor, over $K[x]$.

Thus one obtains the following result :

IX.4.5. *If $f(x)$ be a monic polynomial of degree m, over a perfect field K, and if M be a square matrix, over K, having $(-1)^m . f(x)$ as its characteristic polynomial, then the matrix G, defined in* (2), *page* 203 *has a single elementary divisor, over $K[x]$, given by $[f(x)]^r$.*

It can now be proved that :

IX.4.6. *Any square matrix A, over a perfect field K, is similar to a block-diagonal matrix, $G = diag\ (G_1, \ldots, G_r)$, where $G_1, \ldots, G_r$ are the generalised Jordan blocks corresponding to the elementary divisors of $A - xI$, over $K[x]$.*

Proof: Let G_i be the generalised Jordan block of the form (2), corresponding to on elementary divisor $[f_i(x)]^{r_i}$, where H is taken to be the companion matrix of $f_i(x)$, $f_i(x)$ being a monic polynomial irreducible in $K[x]$. The elementary divisors of $G-xI$, over $K[x]$, are given by the elementary divisors of $G_1-xI,\ldots,G_r-xI$ (by IX.3.5). Again, by IX.3.5, the elementary divisors of $G_1-xI,\ldots,G_r-xI$ constitute the set of elementary divisors of $A-xI$. And, since $A-xI$ and $G-xI$ have the same rank and order, it follows, by the Corollary 2 of IX.1.3, that A is similar to the matrix G.

The matrix G, as defined in IX.4.6, is determined uniquely except for rearrangements of its diagonal blocks ; it is called the *generalised Jordan normal form* (or *Jacobson canonical form*) of the matrix A.

When K is, in particular, an algebraically closed field, the invariant factors are of the form $(x-\lambda)^r$, and the corresponding generalised Jordan block reduces to the (ordinary) Jordan block corresponding to $(x-\lambda)^r$. Consequently, for any matrix A over K, the generalised Jordan normal form becomes identical with the (ordinary) Jordan normal form

If the ground field K be the field of real numbers, then the irreducible monic polynomials over K, are either (i) of the first degree of the form $x-\lambda$, or (ii) of the second degree of the form $f(x)=x^2+bx+c$ (where $b^2<4c$). In case (i), the companion matrix consists of the single element, and the generalised Jordan block corresponding to $(x-\lambda)^r$ is the ordinary Jordan block. In case (ii), the companion matrix M is given by

$$M=\begin{pmatrix} 0 & 1 \\ -c & b \end{pmatrix};$$

and the generalised Jordan block G, coresponding to $[f(x)]^r$, has the form (2). The matrix M above, may be replaced by the matrix

$$M'=\begin{pmatrix} u & v \\ -v & u \end{pmatrix},$$

where $u\pm iv$ (with u, v real) are the roots of $f(x)$, since the characteristic polynomial of M' is also the polynomial $f(x)$.

It is to be noted that the generalised Jordan normal form exists for any square matrix A, over the field of real numbers, while the Jordan normal form of A may not exist.

EXAMPLE 5 : To find the generalised Jordan normal form of the matrix B in example 2 (for which Jordan normal form does not exist).

The invariant factors of $B-xI$ are found (in example 2) to be ;

$$e_1(x)=1,\quad e_2(x)=1,\quad e_3(x)=(x+5)(x^2-3x+3).$$

Thus the elementary divisors of $A-xI$ are

$$x+5 \text{ and } x^2-3x+3.$$

Hence the generalised Jordan normal form of B is the matrix

$$G=\text{diag }(G_1, G_2),$$

where $G_1=(-5)$, and G_2 is the companion matrix of x^2-3x+3. Thus

$$G=\begin{pmatrix} -5 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -3 & 3 \end{pmatrix}.$$

The two roots of x^2-3x+3 are given by $(3\pm i\sqrt{3})/2$. And, one may also take the following matrix H as the normal form, in place of G :

$$H=\begin{pmatrix} -5 & 0 & 0 \\ 0 & 3/2 & \sqrt{3}/2 \\ 0 & -\sqrt{3}/2 & 3/2 \end{pmatrix}.$$

Exercises

1. Write down the Jordan normal form of a matrix, whose Segre characteristic corresponding to the eigen values $\lambda_1, \lambda_2, \lambda_3$ is given by [(2,2,1), (3,1), (2)].

2. Find the Jordan normal forms of the following matrices :

$$\text{(a)}\begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 2 \end{pmatrix},\quad \text{(b)}\begin{pmatrix} 3 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 3 \end{pmatrix},\quad \text{(c)}\begin{pmatrix} 6 & 2 & -1 \\ -1 & 4 & 1 \\ 0 & 1 & 5 \end{pmatrix},\quad \text{(d)}\begin{pmatrix} -2 & -1 & 0 \\ 5 & -1 & 4 \\ 5 & 1 & 2 \end{pmatrix},$$

$$\text{(e)}\begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & -3 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix},\quad \text{(f)}\begin{pmatrix} 2 & 2 & 0 & -2 \\ 0 & 3 & 3 & -2 \\ 0 & -1 & 2 & 1 \\ 3 & 1 & 3 & 0 \end{pmatrix},\quad \text{(g)}\begin{pmatrix} -1 & 1 & 4 & 4 \\ 0 & -1 & 0 & 0 \\ -3 & 1 & 5 & 3 \\ 3 & -1 & -2 & 0 \end{pmatrix}.$$

[*Hints* : The Segre characteristics are (a) [(2,1)], (b) [(1,1),(1)], (c) [(3)], (d) [(2),(1)], (e) [(3),(1)], (f) [(2,1),(1)], and (g) [(2),(1),(1)].

3. Find the possible Jordan normal forms of the real matrix, whose characteristic polynomial is $x^3(1-x^2)$.

4. Let A be a real matrix of order 10, such that the invariant factors of $A-xI$ are given by $e_1(x)=\dots=e_7(x)=1$, $e_8(x)=x-2$, $e_9(x)=(x-2)\,(x^2+x+1)$, and $e_{10}(x)= x-2)^2(x^2+x+1)^2$.

Find the natural normal form, the quasi-natural normal form, and the generalised Jordan normal form of the matrix A, over the real field.

Find also the corresponding normal forms, over the field of complex numbers.

5. For matrices of order 4, over the field of real numbers,
   (i) find all possible Jordan normal forms ;
   (ii) find all possible generalised Jordan normal forms ;
   (iii) find all possible quasi-natural normal forms.

Identify the Jordan normal forms in (i) with the matrices of the collineations of the real projective 3-space, discussed in cases (I)–(IV), pages 178-79.

Find the generalised normal forms in case (ii) above, which correspond to the matrices of the collineations discussed in cases (V)–(VII), page 180, for which Jordan normal forms do not exist.

Find also the normal forms (i), (ii) and (iii), over the field of complex numbers.

5. Unitary similarity.

Let $\Lambda_n(C)$ be the set of all square matrices of order n, over the field of complex numbers C. A matrix B is said to be *unitarily similar* or *unitarily congruent* to a matrix A, if there exists a unitary matrix P, such that $B=PAP^{-1}(=PAP^{*t})$, where $A, B, P \in \Lambda_n(C)$. The relation of being unitarily similar is an equivalence relation, since the unitary matrices form a group. The set $\Lambda_n(C)$ is accordingly partitioned into disjoint classes of unitarily similar matrices.

Hermitian and unitary matrices. A matrix, unitarily similar to a Hermitian or a unitary matrix, is respectively Hermitian or unitary (proof left as an exercise). For the disjoint classes into which the set of Hermitian matrices (or the set of unitary matrices) is partitioned, one obtains the normal forms as follows :

IX.5.1. *Every Hermitian or unitary matrix A is unitarily similar to a diagonal matrix, whose diagonal elements are the eigen values of A.*

Proof : The theorem obviously holds for matrices A of order 1. Let the theorem be true for all matrices of order $< n$, and let A be of order n.

Let X_1 be an eigen vector of A corresponding to an eigen value λ_1, where X_1 may be taken to be a unit vector (since rX_1, for any scalar $r \neq 0$, is also an eigen vector corresponding to λ_1). Let $[X_1, Y_2, \ldots, Y_n]$ be an orthonormal basis of the linear space $V_n(C)$, formed by all n-tuples with complex co-ordinates ; and let Q be the unitary matrix having $X_1, Y_2, \ldots, Y_n$ as its row vectors. Then

$$\begin{aligned}
&\text{the first row-vector of } QAQ^{-1}\\
&\quad=(\text{the first row-vector of } QA).Q^{-1}\\
&\quad=[(\text{the first row-vector of } Q).A].Q^{-1}\\
&\quad=(X_1A).Q^{-1} = (\lambda_1X_1).Q^{-1} = \lambda_1(X_1Q^{-1})\\
&\quad=\lambda_1(\text{the first row-vector of } Q).Q^{-1}\\
&\quad=\lambda_1(\text{the first row-vector of } QQ^{-1})\\
&\quad=\lambda_1(\text{the first row-vector of } I)\\
&\quad=\lambda_1(1,0,\ldots,0)=(\lambda_1,0,\ldots,0).
\end{aligned}$$

Thus
$$QAQ^{-1}=\begin{pmatrix}\lambda_1 & 0\\ L & A_1\end{pmatrix},$$

where A_1 is a square matrix of order $n-1$, and L is a 1-column matrix with $n-1$ elements $a_2, a_3, \ldots, a_n$ (say). Then the first column-vector of QAQ^{-1}, written row-wise, is $(\lambda_1, a_2, \ldots a_n)$. Hence, the first row-vector of $[(QAQ^{-1})^*]^t$ is $(\lambda_1^*, a_2^*, \ldots, a_n^*)$.

If A be Hermitian, then QAQ^{-1} is also Hermitian, and $QAQ^{-1} = [(QAQ^{-1})^*]^t$. Hence $(\lambda_1, 0, \ldots, 0) = (\lambda_1^*, a_2^*, \ldots, a_n^*)$, so that $a_2^* = \ldots = a_n^* = 0$, i.e., $a_2 = \ldots = a_n = 0$.

If, on the other hand, A be unitary, then QAQ^{-1} is also unitary, so that its row-vectors form an orthonormal system. Taking scalar product of the first row-vector, viz. $(\lambda_1, 0, \ldots, 0)$, with the other row-vectors, one obtains $\lambda_1 a_2^* = 0, \ldots, \lambda_1 a_n^* = 0$, whence $a_2^* = \ldots = a_n^* = 0$, i.e., $a_2 = \ldots = a_n = 0$ (since $\lambda_1 \neq 0$, as the first row-vector $(\lambda_1, 0, \ldots, 0)$ is of unit length).

Thus, in either case, $a_2 = \ldots = a_n = 0$, i.e , $L = 0$, so that

$$QAQ^{-1} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & A_1 \end{pmatrix}.$$

As QAQ^{-1} is Hermitian (or unitary), it follows that A_1 is also Hermitian (or unitary, resp.). Now, since the order of A_1 is $n-1$, it follows by the induction hypothesis that there exists a unitary matrix P_1 (of order $n-1$), such that $P_1 A_1 P_1^{-1} = D_1$ is a diagonal matrix. Then

$$R = \begin{pmatrix} 1 & 0 \\ 0 & P_1 \end{pmatrix}$$

is a unitary matrix, and so also if the matrix $P = RQ$. Also

$$\begin{aligned} PAP^{-1} &= (RQ)A(RQ)^{-1} = R(QAQ^{-1})R^{-1} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & P_1 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & A_1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & P_1^{-1} \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 & 0 \\ 0 & P_1 A_1 P_1^{-1} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & D_1 \end{pmatrix} = D \end{aligned}$$

is a diagonal matrix. Thus A is unitarily similar to a diagonal matrix D. The diagonal elements of D are, by IX.2.1, the eigen values of A.

The normal form D, to which a Hermitian or a unitary matrix A can be reduced, by unitary similarity, being uniquely determined by the given matrix A, is therefore canonical.

A more general result for matrices, which are unitarily similar to a diagonal matrix, will now be obtained.

Normal matrices. Let a matrix A be unitarily similar to a diagonal matrix D, so that $D = PAP^{-1} = PAP^{*t}$, where P is a unitary matrix. Then $D^{*t} = (PAP^{*t})^{*t} = PA^{*t}P^{*t} = PA^{*t}P^{-1}$. And from $DD^{*t} = D^{*t}D$, it follows that

$$PAP^{-1} . PA^{*t}P^{-1} = PA^{*t}P^{-1} . PAP^{-1},$$

i.e., $\quad PAA^{*t}P^{-1} = PA^{*t}AP^{-1}$, i.e., $\quad AA^{*t} = A^{*t}A$.

Thus A is a normal matrix. And one obtains :

IX.5.2. *A matrix, which is unitarily similar to a diagonal matrix, is necessarily a normal matrix.*

The converse of IX.5.2 is also true. To prove it, one requires the following lemma :

IX.5.3 *The minimal polynomial of a normal matrix has no repeated linear factors.*

Proof : Let A be a normal matrix, and let $m(x)$ be the minimal polynomial of A. Then $m(x)$ decomposes into linear factors in $C[x]$, where C is the field of complex numbers.

If possible, let $m(x)$ possess a repeated linear factor, say

$$m(x)=g(x)(x-\lambda)^2=f(x).(x-\lambda),$$

where $f(x)=g(x).(x-\lambda)$. Then $g(A).(A-\lambda I)=f(A)\neq O$ (since the degree of $f(x)$ is less than that of $m(x)$. As $f(A)$ is a non-null matrix, there exists a non-null vector Y, such that

$$Y.f(A)=X \text{ is a non-null vector.} \qquad \dots \quad (1)$$

Then $X(A-\lambda I)=Y.f(A).(A-\lambda I)=Y.m(A)=\theta$, since $m(A)$ is the null matrix. Thus X is an eigen vector of the matrix A, corresponding to the eigen value λ. Then, by VIII.5.6, X is an eigen vector of the matrix A^{*t}, corresponding to the eigen value λ^*, so that $X(A^{*t}-\lambda^* I)=\theta$, i.e., $(A-\lambda I)X^{*t}=\theta$ (by taking the conjugate of the transpose). Now,

$$XX^{*t}=Y.f(A).X^{*t}=Y.g(A)(A-\lambda I)X^{*t}=Y.g(A).\theta=0,$$

from which it follows that $X=\theta$, the null vector ; and this contradicts (1).

Hence the minimal polynomial $m(x)$ cannot have repeated linear factors.

It now follows directly from IX.5.3 and IX.2.8 that :

IX.5.4. *A normal matrix A is unitarily similar to a diagonal matrix, whose diagonal elements are the eigen values of A.*

Propositions IX.5.2 and IX.5.4 can now be combined in the form :

IX.5.5. *A complex matrix is unitarily similar to a diagonal matrix if, and only if, it is a normal matrix.*

Since the diagonal matrix D, unitarily similar to a normal matrix A, is uniquely determined by A (as the diagonal elements of D are given by the eigen values of A), it follows that :

IX.5.6. *Two normal matrices are unitarily similar if, and only if, they have the same set of eigen values ; that is, if, and only if, they are similar.*

As every Hermitian, skew-Hermitian or unitary matrix is normal, it follows directly from IX.5.4 and IX.5 6 that :

IX.5.7. (i) *Any Hermitian, skew-Hermitian or unitary matrix A is unitarily similar to a diagonal matrix D, whose diagonal elements are the eigen values of A.*

(*ii*) *Two Hermitian, skew-Hermitian or unitary matrices are unitarily similar if, and only if, they have the same set of eigen values ; that is, if, and only if, they are similar.*

The proposition IX.5.1 is a particular case of the IX.5.7(i). The proof given in IX.5.1 is an independent one and constructive.

Some interesting properties of a normal matrix may now be obtained.

IX.5.8. *If λ be an r-fold eigen value of a normal matrix A of order n, then*

(i) *the rank of $A-\lambda I=n-r$;*

(ii) *the rank of the characteristic sub-space, corresponding to λ, is exactly equal to r ;*

(iii) *there exists a set of r orthonormal eigen vectors of A corresponding to the eigen value λ.*

Proof: The results (i) and (ii) follow directly from IX.2.6, as A is similar to a diagonal matrix.

(iii) The characteristic sub-space, corresponding to λ, being a sub-space of the unitary space $V_n(C)$, is itself a unitary space of rank r. Hence (by II.3.16), it has an orthonormal basis, which consists of r orthonormal eigen vectors of A.

It may be noted that if, for a normal matrix A, $PAP^{-1}=D$ is a diagonal matrix, where P is a unitary matrix, then the successive row-vectors of P are eigen vectors of A corresponding to the successive diagonal elements of D (which are the eigen values of A). If an eigen value λ be repeated r times as a diagonal element of D, then the corresponding row-vectors of P form a set set of r orthonormal eigen vectors of A. The row-vectors of P form an orthonormal basis of $V_n(\mathrm{C})$. And this orthonormal basis is the union of the orthonormal bases of the characteristic sub-spaces, corresponding to the distinct eigen values of the matrix A.

Hermitian form. Let XAX^{*t} be a Hermitian form, so that A is a Hermitian matrix. A substitution $X=YP$ transforms the given Hermitian form to $(YP)A(YP)^{*t}=Y(PAP^{*t})Y^{*t}$. By IX.5.1, there exists a unitary matrix P, such that $PAP^{*t}=D$ is a real diagonal matrix, whose diagonal elements are the eigen values of A. Thus one obtains the following result :

IX.5.9. *Any given Hermitian form* $XAX^{*t}=\Sigma a_{ij}x_ix_j^*$ (*where* $a_{ij}=a_{ji}^*$) *can be reduced to the diagonal normal form,*

$$d_1y_1y_1^*+\ldots+d_ry_ry_r^*,$$

by a substitution $X=YP$, *i.e.,* $x_i=\Sigma p_{ij}y_j$ (*for* $i=1,\ldots,n$), *where* $P=(p_{ij})$ *is a unitary matrix,* $d_1,\ldots,d_r$ *being uniquely determined real numbers, which are the non-zero eigen values of the matrix* A.

EXAMPLE 1 : To reduce the Hermitian form,

$XHX'^t=2x_1x_1+x_2x_2^*+2x_3x_3^*-2ix_1x_2^*+2ix_1^*x_2+(1+i)x_2x_3^*+(1-i)x_2^*x_3$ to its normal form, by a substitution $X=YP$ where P is a unitary matrix; and to find the matrix P.

Soln : The Hermitian matrix H of the Hermitian form is given by

$H=\begin{pmatrix} 2 & -2i & 0 \\ 2i & 1 & 1+i \\ 0 & 1-i & 2 \end{pmatrix}$. Its characteristic polynomial is $x^3+5x^2-2x-8=-(x-4)(x-2)(x+1)$.

The eigen values of H are given by $\lambda=4$, 2, and -1. Hence H is unitarily similar to the diagonal matrix D, having 4,2, and -1 as its diagonal elements. Thus there exists a unitary matrix P, ruch that

$$PHP^{-1}=D=\begin{pmatrix} 4 & & \\ & 2 & \\ & & -1 \end{pmatrix}.$$

The normal form of the Hermitian form is therefore given by

$$4y_1y_1^*+2y_2y_2^*-y_3y_3^*.$$

The row vectors of P are eigen vectors of unit lengths, corresponding to the eigen values $\lambda=4, 2,$ and -1. The eigen vectors, corresponding to an eigen value λ of H are given by the non-null solutions of the system of equations ;

$$\left.\begin{array}{rl} (2-\lambda)x_1+2ix_2 & =0 \\ -2ix_1+(1-\lambda)x_2+(1-i)x_3 & =0 \\ (1+i)x_2+(2-\lambda)x_3 & =0 \end{array}\right\}.$$

For $\lambda=4$, the eigen vectors are given by $r(2i, 2, 1+i)$, for all non-zero complex values of r. The eigen vectors of unit length are obtained by taking $r=\pm 1/\sqrt{10}$.

For $\lambda=2$, the eigen vectors are given by $r(1-i, 0, 2i)$, for all non-zero complex values of r. The eigen vectors of unit length are obtained by taking $r=\pm 1/\sqrt{6}$.

For $\lambda=-1$, the eigen vectors are given by $r(2i, 3, 1+i)$, for all non-zero complex values of r. The eigen vectors of unit length are obtained by taking $r=\pm 1/\sqrt{15}$.

Hence the unitary matrix P is obtained as

$$P=\begin{pmatrix} \frac{2i}{\sqrt{10}} & \frac{2}{\sqrt{10}} & \frac{1+i}{\sqrt{10}} \\ \frac{1-i}{\sqrt{6}} & 0 & \frac{2i}{\sqrt{6}} \\ \frac{2i}{\sqrt{15}} & \frac{3}{\sqrt{15}} & \frac{1+i}{\sqrt{15}} \end{pmatrix}.$$

By taking different sets of signs for the row-vectors, one can obtain 7 such other matrices P.

Exercises

1. Obtain the canonical forms of the following matrices, under unitary similarity :

(a) $\begin{pmatrix} 2 & 3-i \\ 3+i & -1 \end{pmatrix}$, (b) $\begin{pmatrix} \frac{1}{2}+\frac{1}{2}i & \frac{1}{2}-\frac{1}{2}i \\ \frac{1}{2}-\frac{1}{2}i & \frac{1}{2}+\frac{1}{2}i \end{pmatrix}$.

2. Find the unitary matrices which transform the matrices in example 1 to their respective canonical forms.

3. Reduce the following Hermitian forms to their canonical forms under unitary similarity :

(a) $3x_1y^*_1+2x_2y^*_2+(1+i)x_1y_2^*+(1-i)x_2y_1^*$,

(b) $x_1y_1^*+x_2y_2^*+3x_3y_3^*+(1-3i)x_2y_3^*+(1+3i)x_3y_2^*$
$+(2+i)x_1y_3^*+(2-i)x_3y_1^*$.

6. Orthogonal similarity.

Let the ground field be taken, in particular, to be the field of real numbers E. Then the Hermitian, skew-Heritian and unitary matrices reduce to the symmetric, skew-symmetric and orthogonal matrices respectively

A real square matrix B is said to be *orthogonally similar* or *orthogonally congruent* to a matrix A, if there exists an orthogonal matrix P, such that $B=PAP^{-1}(=PAP^t)$.

Real symmetric matrices. Let A be a real symmetric matrix. Since A is necessarily Hermitian there exists a unitary matrix P, such that $PAP^{-1}=D$ is a diagonal matrix. The diagonal elements of D are the eigen values of the symmetric matrix A, and are therefore real numbers (vide VIII.5.8). The row-vectors of P are eigen vectors of A. And since the eigen vectors of the symmetric matrix A have all their co-ordinates real (vide VIII.5.8), it follows that the elements of the unitary matrix P are all real numbers Hence P is an orthogonal matrix ; and A is orthogonally similar to the diagonal matrix D. Thus one obtains :

IX.6.1. *Any real symmetric matrix A is orthogonally similar to a real diagonal matrix D, whose diagonal elements are the eigen values of A.*

A direct proof of IX.6.1 can also be obtained exactly in the same way as in the case of IX.5.1.

Since the normal form D is uniquely determined by the given symmetric matrix A, the normal form is canonical ; and one obtains :

IX.6.2. *Two real symmetric matrices are orthogonally similar if, and only if, they have the same set of eigen values ; that is, if, and only if, they are similar.*

Again, since the eigen vectors of a symmetric matrix have all their co-ordinates real ; that is, all the eigen vectors lie in $V_n(E)$, the characteristic sub-spaces corresponding to the eigen values are sub-spaces of $V_n(E)$. Also, every symmetric matrix is normal. Hence, it follows from IX.5.8 that :

IX.6.3. *If λ be an r-fold eigen value of a real symmetric matrix A of order n, then*

(i) *the rank of $A-\lambda I=n-r$;*

(ii) *the rank of the characteristic sub-space, corresponding to λ, is exactly equal to r ;*

(iii) *there exists a set of r orthonormal eigen vectors of A, corresponding to the eigen value λ.*

Real symmetric bilinear forms. Let XAY^t be a given real symmetric bilinear form, so that A is a symmetric matrix. By a cogradient transformation $X=X'P$ and $Y=Y'P$, the given form is transformed to $X'(PAP^t)Y'^t$. By IX.6.1, there exists an orthogonal matrix P, such that $PAP^t=D$ is a (real) diagonal matrix, whose diagonal elements are the eigen values of A. Hence it follows that :

IX.6.4. *Any real symmetric bilinear form $\Sigma a_{ij}x_iy_j$ (where $a_{ij}=a_{ji}$) can be reduced to the normal (diagonal) form,*

$$d_1x_1'y_1'+\ldots+d_rx_r'y_r',$$

by a suitable cogradient transformation with an orthogonal matrix, where $d_1,\ldots,d_r$ are the non-zero eigen values of the matrix (a_{ij}).

Real quadratic forms. Let XAX^t be a real quadratic form, where A is a symmetric matrix. A substitution $X=YP$ transforms it to the quadratic form $Y(PAP^t)Y^t$. By IX.6.1, there exists an orthogonal matrix P, such that $PAP^t=D$ is a (real) diagonal matrix, whose diagonal elements are the eigen values of A. Hence, it follows that :

IX.6.5. *Any given real quadratic form $XAX^t=\Sigma a_{ij}x_ix_j$ (where $a_{ij}=a_{ji}$) can be reduced to the normal (diagonal) form,*

$$d_1y_1^2+\ldots+d_ry_r^2,$$

by a substitution $X=YP$, i.e., $x_i=\Sigma p_{ij}y_j$ (for $i=1,\ldots,n$) with $P=(p_{ij})$ as an orthogonal matrix, where $d_1,\ldots d_r$ are the non-zero eigen values of the matrix A.

Orthogonal matrices. Since the eigen values of a real orthogonal or a skew-symmetric matrix may not be all real, the matrix may not be orthogonally similar to a diagonal matrix, over the field of real numbers. Hence normal forms of such matrices can only be obtained as block-diagonal matrices. The normal form of a real orthogonal matrix may be obtained as follows :

IX.6.6. *A real orthogonal matrix A of order n is orthogonally similar to a block-diagonal matrix,*

$$M=\text{diag}\,(I_j, -I_m, C_{11}, \,.\,.\,, C_{1r_1}, \ldots, C_{s1}, \ldots, C_{sr_s}),$$

where $C_{uk}=\begin{pmatrix}\cos\theta_u & -\sin\theta_u\\ \sin\theta_u & \cos\theta_u\end{pmatrix}$ *for* $k=1,\ldots, r_u$ *and* $u=1,\ldots,s$. *and* $n=j+m+2r_1+\ldots+2r_s$.

Proof : Let there be j eigen values equal to 1, m eigen values equal to -1, and r_u pairs of eigen values $\cos\theta_u \pm i\,\sin\theta_u$, for $u=1,\ldots,s$, of the given orthogonal matrix A (the non-real eigen values of A, having their moduli equal to 1, occur in conjugate pairs, since the characteristic polynomial of A has real coefficients) ; then $j+m+2r_1+\ldots+2r_s=n$. Since the eigen vectors, corresponding to the real eigen values 1 and -1, have all their co-ordinates real (vide VIII.5.10), and since every real orthogonal matrix is normal, it follows from IX.5.8(iii) that there exist

(i) an orthonormal set of j eigen vectors in $V_n(E)$, $U_1,\ldots,U_j$, satisfying $U_iA=U_i$, for $i=1,\ldots,j$;

(ii) an orthonormal set of m eigen vectors in $V_n(E)$, $V_1,\ldots,V_m$, satisfying $V_iA=-V_i$, for $i=1,\ldots,m$;

(iii) s sets of orthonormal eigen vectors in $V_n(C)$, $\beta_{u1},\ldots,\beta_{ur_u}$, for $u=1,\ldots,s$, satisfying

$$\left.\begin{aligned}\beta_{uk}A &=(\cos\theta_u + i\sin\theta_u)\,\beta_{uk}\\ \beta_{uk}{}^*A &=(\cos\theta_u - i\sin\theta_u)\,\beta_{uk}{}^*\end{aligned}\right\}, \quad \text{for } k=1,\ldots,r_u.$$

Let $\left.\begin{aligned}\beta_{uk}+\beta_{uk}{}^* &= t_{uk}X_{uk}\\ \beta_{uk}-\beta_{uk}{}^* &= it_{uk}Y_{uk}\end{aligned}\right\}$ where $t_{uk}=|\beta_{uk}+\beta_{uk}{}^*|$,

then X_{uk}, Y_{uk} are unit vectors in $V_n(E)$, satisfying

$$\left.\begin{aligned}X_{uk}A &= \cos\theta_u X_{uk} - \sin\theta_{uk}Y_{uk}\\ \text{and}\quad Y_{uk}A &= \sin\theta_u X_{uk} + \cos\theta_u Y_{uk}\end{aligned}\right\}, \quad \begin{aligned}&\text{for } u=1,\,.\,.\,,s,\\ &\text{and } k=1,\ldots r_u.\end{aligned}$$

By VIII.5.11, the vectors in (i), (ii) and (iii) form an orthonormal syetem, so that

$U_i . \beta_{uk}=0$, $U_i . \beta_{uk}^*=0$, $V_i . \beta_{uk}=0$, and $V_i . \beta_{uk}^*=0$; then $U_i . X_{uk}=0$, $U_i . Y_{uk}=0$, $V_i X_{uk}=0$, and $V_i . Y_{uk}=0$,

Hence, it follows that the vectors,

$$U_1, \ldots, U_j\, V_1, \ldots, V_m, X_{11}, Y_{11}, \ldots, X_{1r_1}, Y_{1r_1}, \ldots, X_{sr_s}, Y_{sr_s}, \quad \ldots(1)$$

form an orthonormal set of n vectors in $V_n(\mathrm{E})$.

Let P be the matrix, having the vectors in (1) as its successive row-vectors, then P is a real orthogonal matrix, such that $PAP^{-1}=M$ is the matrix in the required normal form.

Since the normal form M is determined uniquely by the eigen values of A, it follows that :

IX.6.7. *Two real orthogonal matrices are orthogonally similar if, and only if, they have the same set of eigen values; that is, if, and only if, they are similar.*

EXAMPLE 1 : Reduce the following quadratic form to its canonical form, under orthogonal similarity, and find the matrix of the transformation :

$$8x^2+2y^2+2z^2+2xy+2xz+4yz.$$

Soln : The symmetric matrix of the quadratic form is

$$\begin{pmatrix} 3 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 2 \end{pmatrix},$$

and its characteristic polynomial is $-x^3+7x^2-10x=-x(x-2)(x-5)$; the eigen values are $\lambda=5$, 2 and 0.

Hence the canonical form is given by $5X^2+2Y^2$.

The successive row-vectors of the orthogonal transformation, which transforms the given quadratic form into the canonical form, are the unit eigen vectors of the matrix corresponding to the eigen values 5, 2 and 0 respectively.

The eigen vectors, corresponding to an eigen value λ, are given by the non-null solutions of the system of equations,

$$\left.\begin{array}{llll} (8-\lambda)x_1 + & x_2 & + \quad x_3 & =0 \\ x_1 & + (2-\lambda)x_2 & + \quad 2x_3 & =0 \\ x_1 & + \quad 2x_2 & + (2-\lambda)x_3 & =0 \end{array}\right\}.$$

For $\lambda=5$, the unit eigen vectors are $\pm(1/\sqrt{3},\ 1/\sqrt{8},\ 1/\sqrt{3})$,
for $\lambda=2$, the unit eigen vectors are $\pm(-2\sqrt{6},\ 1/\sqrt{6},\ 1/\sqrt{6})$,
and for $\lambda=0$, the unit eigen vectors are $\pm(0,\ -1/\sqrt{2},\ 1/\sqrt{2})$.

Hence the required orthogonal matrix is

$$\mathrm{P}=\begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{8}} \\ \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{-1}{\sqrt{2}} & \frac{2}{\sqrt{2}} \end{pmatrix}.$$

[For different choice of the signs of the row-vectors, one may also obtain 7 other matrices, each of which effects the required transformation].

EXAMPLE 2 : Find the canonical form to which the following orthogonal matrix can be reduced under orthogonal similarity :

$$\begin{pmatrix} \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \end{pmatrix}.$$

The characteristic polynomial of the given matrix is

$$-x^3+\tfrac{2}{3}x^2+\tfrac{2}{3}x-1=-(x+1)(x^2-\tfrac{5}{3}x+1) \;;$$

and the eigen values are

$$-1 \text{ and } \frac{5}{6}\pm i\frac{\sqrt{11}}{6}=\cos\theta\pm i\sin\theta \text{ (say).}$$

Hence the required canonical form is given by

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}=\begin{pmatrix} -1 & 0 & 0 \\ 0 & \frac{5}{6} & -\frac{\sqrt{11}}{6} \\ 0 & \frac{\sqrt{11}}{6} & \frac{5}{6} \end{pmatrix}.$$

Exercises

1. Obtain the canonical forms, under orthogonal smilarity, for the following quadratic forms :

(a) $x^2+5y^2+z^2+2xy+6xz+2yz$,

(b) $2x^2+z^2+2xy+3xz+yz$,

(c) $10x^2-7y^2+z^2-8xy+8xz-4yz$,

(d) $x^2+3y^2+z^2-3w^2+4yw-10xz$.

2. Find the canonical forms of the following symmetric bilinear forms, under orthogonal similarity :

(a) $2x_1y_2+2x_2y_1+3x_2y_2$;

(b) $x_1y_1+2x_1y_2+2x_2y_1+4x_2y_2$;

(c) $5x_1y_1+3x_2y_2+3x_3y_3+x_1y_2+x_2y_1-x_1y_3-x_3y_1$
$-x_2y_3-x_3y_2$;

(d) $x_1y_1+4x_2y_2+4x_3y_3+2x_1y_2+2x_2y_1-2x_1y_3-2x_3y_1$
$-4x_2y_3-4x_3y_2$.

3. Reduce the following quadratic forms to their canonical forms, under orthogonal similarity, and find the matrix of the transformation in each case :

(a) $5x^2+6y^2+7z^2-4xy-4yz$,

(b) $x^2-2y^2+z^2-4xy-8xz+4yz$,

(c) $8x^2+4xy-4xz-2yz$.

[*Hints* : In (b), the three eigen values are $6, -3$ and -3. The unit eigen vectors corresponding to 6 are $\pm(2/3, -1/3, -2/3)$. The characteristic sub-space, corresponding to -3, is of rank 2, and consists of the solutions of the equation $2x_1-x_2-2x_3=0$. A pair of orthonormal eigen vectors, corresponding to -3, i.e., an orthonormal pair of solutions of the above equation, may be chosen as $\pm(2/3, 2/3, 1/3)$ and $\pm(1/3, -2/3, 2/3)$. The matrix of the orthogonal transformation can be formed with the above eigen vectors as its row vectors]

4 Find the canonical forms for real orthogonal matrices of (i) order 3, and (ii) order 4, under orthogonal similarity.

6. 1. Application (Classification of quadrics in the Euclidean space).

An equation of the second degree in n variables of the form

$$\Sigma a_{ij}x_ix_j + \Sigma a_ix_i + a = 0 \qquad \ldots (1)$$

is said to represent a hyperquadric in the Euclidean n-space. In matrix notation, it is represented by

$$XAX^t + A'X^t + a = 0, \qquad \ldots (2)$$

where A and A' are respectively a square matrix of order n and a 1-row matrix (formed by the coefficients of the homogeneous linear part of the given equation). Without any loss of generality, the square matrix A may be taken to be symmetric. By an orthogonal transformation $X=YP$, the equation (2) is transformed to

$$(YP)A(YP)^t + A'(YP)^t + a = 0,$$

i.e.,
$$Y(PAP^t)Y^t + (A'P^t)Y^t + a = 0. \qquad \ldots (3)$$

By IX.6.5, there exists an orthogonal matrix P, such that $PAP^t=D$ is a diagonal matrix, whose diagonal elemets $d_1,\dots,d_n$ are the eigen values of A. Hence, by the corresponding orthogonal transformation $X=YP$, the equation (3) reduces to

$$d_1y_1^2+\dots+d_ry_r^2+a_1'y_1+\dots+a_n'y_n+a=0, \qquad \dots (4)$$

where r is the rank of the matrix A (so that $n-r$ eigen values of A are 0). Then the translation (parallel displacement),

$$\left.\begin{array}{ll} y_i=y_i'-a_i'/2d_i, & \text{for } i=1,\dots,r, \\ \text{and} \quad y_i=y_i', & \text{for } i=r+1,\dots n \end{array}\right\},$$

transforms the equation (4) to

$$d_1y_1'^2+\dots+d_ry_r'^2+a_{r+1}'y_{r+1}'+\dots+a_n'y_n+a'=0. \qquad \dots (5)$$

Two cases may now arise :

(I) Let $a_{r+1}'=\dots=a_n'=0$; then (5) reduces to

$$d_1y_1'^2+\dots+d_ry_r'^2+a'=0. \qquad \dots (6)$$

(II) Let one at least of the coefficients $a_{r+1}',\dots,a_n'$ be non-zero, say $a_{r+1}'\neq 0$. Let M be an orthogonal matrix of order $n-r$, having $a_{r+1}'/u,\dots,a_n'/u$ as the successive elements in its first column (the existence of such a matrix is assured by II.3.15), where $u=\sqrt{(a_{r+1}'^2+\dots+a_n'^2)}$. Then the orthogonal transformation, defined by

$$\left.\begin{array}{ll} w_i=y_i', & \text{for } i=1,\dots,r, \\ \text{and} \quad (w_{r+1},\dots,w_n)=(y_{r+1}',\dots,y_n').M & \end{array}\right\},$$

transforms the equation (5) to the equation

$$d_1w_1^2+\dots+d_rw_r^2+uw_{r+1}+a'=0. \qquad \dots (7)$$

Next, the translation,

$$\left.\begin{array}{l} w_i \;\;\;\, =X_i, \text{ for } i=1,\dots,r,r+2,\dots,n \\ \text{and} \quad w_{r+1}=X_{r+1}-a'/u \end{array}\right\}$$

transforms the equation (7) to the equation

$$d_1X_1^2+\dots+d_rX_r^2+uX_{r+1}=0. \qquad \dots (8)$$

Thus, the given equation (1) reduces to one of the equations (6) and (8), by a suitable Euclidean motion. The quadratic part, in either case, is completely determined by the eigen values of the matrix A. The case for the equation (8), i.e., case (II), arises only when $r < n$.

Putting $d_i/a' = -c_i$, when $a' \neq 0$, and $d_i = c_i$, when $a' = 0$, and changing y_i' to X_i, one obtains the following two types from the equation (6) :

$$c_1 X_1^2 + \ldots + c_r X_r^2 = 1, \qquad \ldots (9)$$

and

$$c_1 X_1^2 + \ldots + c_r X_r^2 = 0. \qquad \ldots (10)$$

Putting $d_i/u = -c_i$, for $i = 1, \ldots, r$, the equation (8) assumes the form

$$c_1 X_1^2 + \ldots + c_r X_r^2 = X_{r+1}. \qquad \ldots (11)$$

Thus one obtains :

IX.6.8. *The equation of any given hyperquadric in the Euclidean n-space can always be reduced to one of the canonical forms* (9), (10) *and* (11) *by a suitable Euclidean motion.*

Hyperquadrics represented by equations (9) and (10) are called *central*, of which those corresponding to (10) are distinguished as *conical* ; hyperquadrics represented by equations (11) are called *parabolic*. If $r < n$ in (9) or (10), and $r < n-1$ in (11), the hyperquadric is said to be *cylindrical*.

These classes are divided into further sub-classes, according to the different values of the numerical invariants r and s, the rank and signature of the homogeneous quadratic part of the equations. To distinguish the positive coefficients from the negative ones, one puts $|c_i| = 1/b_i^2$, for $i = 1, \ldots, r$. Further, the signature is usually taken to be non-negative ; for this purpose, it may be necessary to change the signs of the terms on the right hand side of the equations (9), (10) and (11). By such a change, the equations (10) and (11) remain unaffected, while in case (9), 1 may have to be changed to -1.

Thus the equation (9), (10), (11) can be put in the forms :

$$\frac{X_1^2}{b_1^2} + \ldots + \frac{X_p^2}{b_p^2} - \frac{X_{p+1}^2}{b_{p+1}^2} - \ldots - \frac{X_r^2}{b_r^2} = 1 \qquad \ldots (12)$$

$$= -1 \qquad \ldots (13)$$

$$= 0 \qquad \ldots (14)$$

$$= X_{r+1}, \qquad \ldots (15)$$

where $0 \leqslant r-p \leqslant p \leqslant r \leqslant n$, so that $0 \leqslant 2p - r = s \leqslant r \leqslant n$.

Conics in Euclidean plane. For the Euclidean plane, $n=2$, and the hyperquadrics are called conics. The normal forms of different classes of conics are obtained from the equations (12)–(15), corresponding to different possible values of the numerical invariants r and s, subject to the condition $0 \leq s \leq r \leq 2$, as follows (where X_1, X_2 are replaced by x, y, and b_1, b_2 by a, b respectively) :

	r	*s*	*Normal form*	*Description*
(i)	2	2	$\frac{x^2}{a^2}+\frac{y^2}{b^2}= 1$,	Ellipse.
(ii)			$=-1$,	Without any real trace.
(iii)			$= 0$,	Point conic, (0, 0).
(iv)	2	0	$\frac{x^2}{a^2}-\frac{y^2}{b^2}= 1$ (or -1),	Hyperbola.
(v)			$= 0$,	A pair of intersecting straight lines, $x/a \pm y/b=0$.
(vi)	1	1	$x^2/a^2= 1$,	A pair of parallel straight lines, $x=\pm a$.
(vii)			$=-1$,	Without any real trace.
(viii)			$= 0$,	A pair of coincident straight lines, $x=0$.
(ix)			$= y$,	Parabola.

Quadrics in Euclidean 3-space. For the Euclidean 3-space, $n=3$, and the hyperquadrics are called quadrics. The normal forms of the different classes of quadrics are obtained from the equations (12)–(15), corresponding to different values of r and s, where $0 \leq s \leq r \leq 3$, as follows (X_1, X_2, X_3 being replaced by x, y, z, and b_1, b_2, b_3 by a, b, c respectively) :

	r	*s*	*Normal forms*	*Description*
(i)	3	3	$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}= 1$,	Ellipsoid.
(ii)			$=-1$,	Without any real trace.
(iii)			$= 0$,	Point quadric, (0,0, 0).

(iv)	3	1	$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}$	$= 1$,	Hyperboloid of one sheet.
(v)				$=-1$,	Hyperboloid of two sheets.
(vi)				$= 0$,	Quadric cone.
(vii)	2	2	$\frac{x^2}{a^2}+\frac{y^2}{b^2}$	$= 1$,	Elliptic cylinder.
(viii)				$=-1$,	Without any real trace.
(ix)				$= 0$,	Line quadric, $x=0$, $y=0$.
(x)				$= z$,	Elliptic paraboloid.
(xi)	2	0	$\frac{x^2}{a^2}-\frac{y^2}{b^2}$	$=1$ (or -1),	Hyperbolic cylinder.
(xii)				$= 0$,	A pair of planes intersecting along the z-axis
(xiii)				$= z$,	Hyperbolic paraboloid.
(xiv)	1	1	$\frac{x^2}{a^2}$	$= 1$,	A pair of parallel planes, $x=\pm a$.
(xv)				$=-1$,	Without any real trace.
(xvi)				$= 0$,	A pair of coincident planes, $x=0$.
(xvii)				$= y$,	Parabolic cylinder.

The 17 different types of quadrics, described above, determine the 17 distinct *affine classes* into which the quadrics can be divided. Since the coefficients in the above normal forms are invariants under Euclidean motions, distinct *Euclidean classes* are obtained for distinct sets of coefficients. Thus each of the above affine classes (excepting (xvi) yields infinitely many Euclidean classes, corresponding to different sets of values for the coefficients a, b, c.

Normal forms of the 17 distinct affine classes *in affine coordinates* can be deduced from the above normal forms, by the transformations :

$x=aX$, $y=bY$, $z=cZ$ for the types (i)–(vi),

$x=aX$, $y=bY$, $z=Z$ for the types (vii)–(xiii),

and $x=aX$, $y=Y$, $z=Z$ for the types (xiv)–(xvii).

The affine normal forms, corresponding to the same values of r and s and having the same descriptions, as above, are given respectively by :

(i)	$X^2+Y^2+Z^2=1$	(x)	$X^2+Y^2=Z$
(ii)	$=-1$	(xi)	$X^2-Y^2=1$ (or -1)
(iii)	$=0$	(xii)	$=0$
(iv)	$X^2+Y^2-Z^2=1$	(xiii)	$=Z$
(v)	$=-1$	(xiv)	$X^2=1$
(vi)	$=0$	(xv)	$=-1$
(vii)	$X^2+Y^2=1$	(xvi)	$=0$
(viii)	$=-1$	(xvii)	$=Y.$
(ix)	$=0$		

The affine normal forms may also be obtained directly from the given equation. The matrix of the transformation is required to be non-singular, but not necessarily orthogonal. Hence, at the first step, a non-singular matrix P is found, for which PAP^t is a diagonal matrix, having $1, -1, 0$ as its diagonal elements. Then the first degree terms in those variables, for which the square terms are present, are abosorbed in the square terms by completing squares (i. e., by a suitable parallel displacement). For the remaining linear portion, a non-singular transformation (which need not be orthogonal) can readily be found, by which the equation is transformed to the affine normal form.

EXAMPLE 1 : Find the canonical form of the equation for the following quadric, under Euclidean motion of the Euclidean 3-space ; find also the Euclidean motion, by which the given equation is reduced to the canonical form :

$$3x^2+2y^2+2z^2+2xy+2xz+4yz-2x+6y-4z+3=0.$$

Soln : It has been shown in example 1, page 218, that the homogeneous quadratic part, $3x^2+2y^2+2z^2+2xy+2xz+4yz$ of the given equation can be reduced to the canonical diagonal form $5X^2+2Y^2$, by the orthogonal transformation,

$$(x,y,z)=(X,Y,Z)\begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}, \quad \text{i.e.,} \quad \begin{aligned} x&=\frac{1}{\sqrt{3}}X-\frac{2}{\sqrt{6}}Y \\ y&=\frac{1}{\sqrt{3}}X+\frac{1}{\sqrt{6}}Y-\frac{1}{\sqrt{2}}Z \\ z&=\frac{1}{\sqrt{3}}X+\frac{1}{\sqrt{6}}Y+\frac{1}{\sqrt{2}}Z. \end{aligned}$$

The linear part is thereby transformed into

$$\begin{aligned} -2x+6y-4z+3&=-2\left(\frac{1}{\sqrt{3}}X-\frac{2}{\sqrt{6}}Y\right)+6\left(\frac{1}{\sqrt{3}}X+\frac{1}{\sqrt{6}}Y-\frac{1}{\sqrt{2}}Z\right) \\ &\qquad -4\left(\frac{1}{\sqrt{3}}X+\frac{1}{\sqrt{6}}Y+\frac{1}{\sqrt{2}}Z\right)+3 \\ &=\sqrt{6}Y-5\sqrt{2}Z+3. \end{aligned}$$

Thus the given equation, as a whole, is transformed into

$$5X^2+2Y^2+\sqrt{6}Y-5\sqrt{2}Z+3=0$$

i.e.,

$$5X^2+2\left(Y+\frac{\sqrt{6}}{4}\right)^2-5\sqrt{2}\left(Z-\frac{9\sqrt{2}}{40}\right)=0,$$

which is transformed by the translation,

$$X=x', \quad Y=y'-\frac{\sqrt{6}}{4}, \quad Z=z'+\frac{9\sqrt{2}}{40},$$

into the canonical form

$$5x'^2+2y'^2=5\sqrt{2}z' \quad \text{i.e.,} \quad \frac{x'^2}{\sqrt{2}}+\frac{y'^2}{5/\sqrt{2}}=z'.$$

The quadric is therefore an elliptic paraboloid.

The Euclidean motion, which transforms the given equation to the canonical form, has the equations :

$$x=\frac{1}{\sqrt{3}}X-\frac{2}{\sqrt{6}}Y \qquad =\frac{1}{\sqrt{3}}x'-\frac{2}{\sqrt{6}}\left(y'-\frac{\sqrt{6}}{4}\right)$$

$$y=\frac{1}{\sqrt{3}}X+\frac{1}{\sqrt{6}}Y-\frac{1}{\sqrt{2}}Z \quad =\frac{1}{\sqrt{3}}x'+\frac{1}{\sqrt{6}}\left(y'-\frac{\sqrt{6}}{4}\right)-\frac{1}{\sqrt{2}}\left(z'+\frac{9\sqrt{2}}{40}\right)$$

$$z=\frac{1}{\sqrt{3}}X+\frac{1}{\sqrt{6}}+\frac{1}{\sqrt{2}}Z \quad =\frac{1}{\sqrt{3}}x'+\frac{1}{\sqrt{6}}\left(y'-\frac{\sqrt{6}}{4}\right)+\frac{1}{\sqrt{2}}\left(z'+\frac{9\sqrt{2}}{40}\right)$$

i.e.,

$$\left.\begin{aligned} x&=\frac{1}{\sqrt{3}}x'-\frac{2}{\sqrt{6}}y' \qquad\qquad +\frac{1}{2} \\ y&=\frac{1}{\sqrt{3}}x'+\frac{1}{\sqrt{6}}y'-\frac{1}{\sqrt{2}}z'-\frac{19}{40} \\ z&=\frac{1}{\sqrt{3}}x'+\frac{1}{\sqrt{6}}y'+\frac{1}{\sqrt{2}}z'-\frac{1}{40} \end{aligned}\right\}.$$

EXAMPLE 2 : Determine the type of the following quadric in the Euclidean 3-space : $3x^2+y^2-2xy+2xz+6x+2z+9=0$.

Soln : Let A and B denote respectively the matrices,

$$A=\begin{pmatrix} 3 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad B=\begin{pmatrix} 3 & -1 & 1 & 3 \\ -1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 3 & 0 & 1 & 9 \end{pmatrix},$$

where A is the symmetric matrix of the homogeneous quadratic part, and B the symmetric matrix formed with coefficients of all the terms in the given equation. Then det (A) and det (B) are invariants, under Euclidean motions.

The characteristic polynomial of A is

$$\psi_A(x)=-x^3+4x^2+x-1$$

whose roots may not be rational, and cannot be determined by simple factorisation. In this case the type of the quadric may be determined as follows.

Since $\det(A) \neq 0$, the rank of the quadratic part is 3. Thus the only possible cases are types (i)-(vi).

$\text{Det}(A) = -1 \neq 0$ implies that the quadric cannot be of types (iii) and (vi). Again, since $\det(B) = -5$, i.e., negative, the types (ii) and (iv) are eliminated.

The product of the eigen values of A is -1, i.e., negative, while the sum of the eigen values is 4, i.e., positive; hence the type (i) is not possible.

Hence the quadric must be of type (v); that is, the quadric is a hyperboloid of two sheets.

Exercises

1. Determine the affine normal forms of the following conics:

(a) $2x^2 - 2xy + 5y^2 - 2x - 8y + 4 = 0$;
(b) $x^2 - 3xy + y^2 + 10x - 10y + 21 = 0$;
(c) $x^2 - 2xy + y^2 + 3x + 2y + 2 = 0$;
(d) $5x^2 - 2xy + y^2 - 8(x + y - 1) = 0$.

2 Find the affine transformations which transform the equations (a)–(d) in exercise 1 to their respective affine normal forms.

3. (*Pembroke's example*) Determine the different types of conics represented by the following equation, as λ varies from a large negative to a large positive value:

$$x^2 + 4\lambda xy + 4y^2 + 2(1+\lambda)x + 8y + (5 + 2\lambda) = 0.$$

4. Determine the types of the following quadrics in the Euclidean 3-space:

(a) $x^2 + 5y^2 + z^2 + 2xy + 6xz + 2yz = 5$;
(b) $2x^2 + z^2 + 2xy + 3xz + yz - 2 = 0$;
(c) $x^2 - 2y^2 + z^2 - 4xy - 8xz + 4yz + 6 = 0$;
(d) $10x^2 - 7y^2 + z^2 - 8xy + 8xz - 8yz = 0$;
(e) $5x^2 + 6y^2 + 7z^2 - 4xy - 4yz + 18x - 24y - 6z + 18 = 0$;
(f) $8x^2 + 4xy - 4xz - 2yz + 2x - 6y + 2z - 20 = 0$.

5. Find the Euclidean motions which transform the equations (a)–(f) in exercise 4 to their respective canonical forms.

6. (*Modenov method*) : Show that for the hyperquadric with equation (1), the values of a' in (6) and u in (8) are given by

$$a' = v_{r+1}/(d_1 \ldots d_r) \text{ and } u = 2\sqrt{[-v_{r+2}/(d_1 \ldots d_r)]},$$

where $v_1, \ldots, v_{n+1}$ are given by

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & \ldots & a_{1n} & a_1/2 \\ a_{21} & a_{22}-\lambda & \ldots & a_{2n} & a_2/2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn}-\lambda & a_n/2 \\ a_1/2 & a_2/2 & \ldots & a_n/2 & a \end{vmatrix} = v_{n+1} - v_n\lambda + v_{n-1}\lambda^2 - \ldots + v_1(-\lambda)^n.$$

[See Shilov, G.E. – An Introduction to the theory of Linear Spaces, pp 236-243].

7. Simultaneous reduction of a pair of real symmetric matrices.

Let A and B be a pair of real symmetric matrices of order n, of which the eigen values of A are positive. There exists an orthogonal matrix P, such that $PAP^t = D$ is a diagonal matrix, whose diagonal elements $d_1, \ldots, d_n$ are all positive real numbers. Let Q be the diagonal matrix, whose diagonal elements are $1/\sqrt{d_1}, \ldots, 1/\sqrt{d_n}$. Then $QDQ^t = I$, the identity matrix. Thus

$$(QP)A(QP)^t = Q(PAP^t)Q^t = QDQ^t = I.$$

Again $(QP)B(QP)^t$ is a symmetric matrix, since B is symmetric. Correspondingly, there exists an orthogonal matrix R, such that $R[(QP)B(QP)^t]R^t = G$ is a diagonal matrix, whose diagonal elements $g_1, \ldots, g_n$ are the eigen values of the (symmetric) matrix $(QP)B(QP)^t$. Let $H = RQP$; then H is a non-singular matrix, since each of R, Q and P is so. Also

$$HBH^t = (RQP)B(RQP)^t = R[(QP)B(QP)^t]R^t = G \quad \ldots (1)$$

and $HAH^t = (RQP)A(RQP)^t = R[(QP)A(QP)^t]R^t = RIR^t = I \ldots (2)$

(since R is an orthogonal matrix). Again, since for arbitrary λ,

$$H(B-\lambda A)H^t = HBH^t - \lambda HAH^t = G - \lambda I$$

$$= \text{diag}(g_1 - \lambda, \ldots, g_n - \lambda),$$

it follows that $g_1, \ldots, g_n$ are the roots of $\det(B - xA) = 0$. Thus

IX.7.1. *If A and B be a pair of real symmetric matrices of the same order. of which the eigen values of A are all positive then there exists a non-singular matrix H, such that $HAH^t = I$, the identity matrix, and $HBH^t = G$ is a diagonal matrix, whose diagonal elements are the roots of* $\det(B - xA) = 0$.

Reduction of a pair of symmetric bilinear forms. Let XAY^t and XBY^t be a pair of real symmetric bilinear forms. Then a cogredient transformation $X = X'H$, $Y = Y'H$ transforms the given forms into the forms $X'(HAH^t)Y'^t$ and $X'(HBH^t)Y'^t$ respectively. Hence, it follows directly from IX.7.1 that :

IX.7.2. *Any given pair of real symmetric bilinear forms XAY^t and XBY^t, of which the eigen values of A are positive, can be transformed, by a suitable non-singular transformation, respectively to the canonical forms,*

$$X'Y'^t = x_1'y_1' + \dots + x_n'y_n' \text{ and } X'GY'^t = g_1x_1'y_1' + \dots + g_nx_n'y_n'$$

where $g_1, \dots, g_n$ are the roots of $\det(B - xA) = 0$.

Reduction of a pair of real quadratic froms. Let XAX^t and XBX^t be a pair of real quadratic forms, with symmetric matrices A and B. Then a transformation $X = YH$ transforms the two forms into the forms $Y(HAH^t)Y^t$ and $Y(HBH^t)Y^t$ respectively. Hence, it follows from IX.7.1 that :

IX.7.3. *If XAX^t and XBX^t be a pair of real quadratic forms with symmetric matrices A, B of order n, of which XAX^t is positive definite, then there exists a non-singular transformation $X = YH$, which transforms the two given forms respectively to the canonical forms,*

$$YY^t = y_1^2 + \dots + y_n^2 \text{ and } YGY^t = g_1y_1^2 + \dots + g_ny_n^2,$$

where $g_1, \dots, g_n$ are the roots of $\det(B - xA) = 0$.

EXAMPLE 1 : Let $x^2 + 2xy + 2y^2$ and $3x^2 + 6xy - y^2$ be a given pair of quadratic forms, the first of which is positive definite. To reduce them simultaneously to their canonical forms.

Soln : The symmetric matrices of the two forms are

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 3 & 3 \\ 3 & -1 \end{pmatrix}, \qquad \text{Then } B - xA = \begin{pmatrix} 3-x & 2 \\ 2 & -1-2x \end{pmatrix}.$$

so that $\det(B-xA)=(3-x)(-1-2x)-4=(2x-7)(x+1)$. Hence the two given forms can be reduced to their canonical forms, X^2+Y^2 and $-X^2+\frac{7}{2}Y^2$ simultaneously, by a suitable non-singular transformation.

Exercises

1. Reduce the following pairs of real quadratic forms simultaneously, the first one to a sum of squares and the second one to the canonical diagonal form, by suitable nonsingular transformations :

(a) $x^2+2xy+3y^2$, $2x^2+4xy+4y^2$;

(b) $2x^2+5y^2+5z^2+4xy+4xz+8yz$,

$7x^2+y^2+z^2-8xy-8xz+16yz$.

2. Find the non-singular matrices which transform the given pairs of quadratic forms in example 1 to their canonical forms.

3. Use examples 1(b) and (2) to find a common self-polar triangle for the following pair of conics in the real projective plane

$$2x_1^2+5x_2^2+5x_3^2+4x_1x_2+4x_1x_3+8x_2x_3=0,$$

and $$7x_1^2+x_2^2+x_3^2-8x_1x_2-8x_1x_3-16x_2x_3=0.$$

4. Generalise the proposition IX.7.3 to the following :

Each pair of Hermitian froms, where the first one is positive definite, can be reduced to the canonical forms,

$$z_1z_1^*+\ldots+z_nz_n^* \text{ and } g_1z_1z_1^*+\ldots+g_nz_nz_n^*,$$

by a suitable non-singular transformation, the coefficients g_1, ..., g_n being determined uniquely by the two given forms.

CHAPTER X

LINEAR OPERATORS

A single-valued mapping of a set R into itself has been called an *operator* on R, or an *endo-mapping* or a *transformation of R* (vide page 6).

Let $\mathbf{A}$ be an operator on R, and let P be a non-empty subset of R. Then $P\mathbf{A}$, called the *image of P under* $\mathbf{A}$, is the subset of R, formed by all $x\mathbf{A}$, as x runs over P. And $P\mathbf{A}^{-1}$, called the *inverse image of P* or the *pre-image of P under* $\mathbf{A}$, is the subset of R, consisting of all those elements y, such that $y\mathbf{A} \varepsilon P$. The pre-image of P may as well be empty, unless $\mathbf{A}$ is surjective. Taking P to be a one-pointic subset, consisting of the point x, say, the pre-image $x\mathbf{A}^{-1}$ is always a nonempty subset of R, provided that $\mathbf{A}$ is surjective. Also, $x\mathbf{A}^{-1}$ consists of one point only, for every point x, provided that $\mathbf{A}$ is injective. Thus $\mathbf{A}^{-1}$ is also an operator on R if, and only if, $\mathbf{A}$ is bijective, and then $\mathbf{A}^{-1}$ is also bijective. In this case, the operator $\mathbf{A}$ is said to be *invertible*. Thus, *an operator* $\mathbf{A}$ *is invertible if, and only if, it is a bijective mapping of R onto R.*

1. Linear operators on a linear space.

Let V be a linear space, over a field K. If $\mathbf{A}$ be an operator on V, such that

$$(a\alpha+b\beta)\mathbf{A}=a(\alpha\mathbf{A})+b(\beta\mathbf{A}), \qquad \text{....} \ (1)$$

where a, b are arbitrary scalars from K, and α, β are any two vectors in V, then $\mathbf{A}$ is called a *linear operator on the linear space V.*

The condition (1) may as well be replaced by a pair of equivalent conditions, by taking (i) $a=b=1$, and (ii) $b=0$,

(i) $(\alpha+\beta)\mathbf{A}=\alpha\mathbf{A}+\beta\mathbf{A}$, and (ii) $(a\alpha)\mathbf{A}=a(\alpha\mathbf{A})$, ... (2)

where α, β are arbitrary vectors in V, and a is any scalar from K.

The operator O, which maps every vector α into the null vector θ, i.e., for which $\alpha O=\theta$, for every $\alpha \in V$, is obviously a linear operator on V; it is called the *null operator* on V. Likewise, the *identity operator* I is the linear operator on V, distinguished by the property, that $\alpha I=\alpha$, for every $\alpha \in V$.

Besides these two trivial linear operators, there exsit others ; in fact,

P. **X.1.1.** *If $\alpha_1,\dots,\alpha_n$ be a basis of a linear space V, over a field K, and if $\beta_1,\dots,\beta_n$ be any given set of n vectors in V, then there exists a unique linear operator on V which transforms the vectors $\alpha_1,\dots,\alpha_n$ respectively to the vectors $\beta_1,\dots,\beta_n$.*

Proof : Let ξ be an arbitrary vector belonging to V. Then ξ can be expressed uniquely as a linear combination,

$$\xi=x_1\alpha_1+\dots+x_n\alpha_n. \qquad \dots (3)$$

Let A be an operator on V, defined by

A=T

$$\xi A=x_1\beta_1+\dots+x_n\beta_n.$$

Then A is a linear operator on V. In fact, if a be any scalar, and $\eta=y_1\alpha_1+\dots+y_1\alpha_n$ be any vector in V, then

$$\begin{aligned}(\xi+\eta)\,A&=[(x_1+y_1)\,\alpha_1+\dots+(x_n+y_n)\,\alpha_n]\,A\\&=(x_1\alpha_1+\dots+x_n\alpha_n)A+(y_1\alpha_1+\dots+y_n\alpha_n)\,A\\&=\xi A+\eta A,\end{aligned}$$

and

$$\begin{aligned}(a\xi)\,A&=[(ax_1)\,\alpha_1+\dots+(ax_n)\,\alpha_n\,]\,A\\&=[a(x_1\alpha_1+\dots+x_n\alpha_n)]\,A\\&=a[(x_1\alpha_1+\dots+x_n\alpha_n)A]=a(\xi A).\end{aligned}$$

It can be readily seen, by taking $x_1=1$, $x_2=0,\dots,x_n=0$, that $\alpha_1 A=\beta_1$; and similarly, $\alpha_2 A=\beta_2,\dots,\alpha_n A=\beta_n$. Thus there exists a linear operator A on V, which transforms α_i to β_i, for $i=1,2,\dots,n$.

To prove the uniqueness of the linear operator A, let there exist a linear operator B, such that $\alpha_i B=\beta_i$, for $i=1,2,\dots,n$. Then, for an arbitrary vector ξ, given by (3), one gets

$$\begin{aligned}\xi B&=(x_1\alpha_1+\dots+x_n\alpha_n)\,B=x_1(\alpha_1 B)+\dots+x_n(\alpha_n B)\\&=x_1\beta_1+\dots+x_n\beta_n=\xi A\,;\end{aligned}$$

consequently, B = A.

Matrix representation. The above proposition can conveniently be used to find a matrix representation for a linear operator.

X.12 *A linear opeator* $\mathbf{A}$ *on a linear space* V *of rank* n, *over a field* K, *is represented by an* $n \times n$ *matrix* A, *over* K, *with respect to a fixed basis of* V.

Proof: Let $\alpha_1, \ldots, \alpha_n$ be a given basis of V, and let

$$\alpha_i \mathbf{A} = \alpha_1 a_{i1} + \ldots + \alpha_n a_{in}, \text{ for } i = 1, 2, \ldots n.$$

Let ξ be an arbitray vector belonging to V, and let ξ and $\xi\mathbf{A}$ have respectively the co-ordinate vectors $X = (x_1, \ldots, x_n)$ and $X' = (x_1', \ldots, x_n')$, when referred to the basis $\alpha_1, \ldots, \alpha_n$; then

$$\xi = x_1\alpha_1 + \ldots + x_n\alpha_n, \text{ and } \xi\mathbf{A} = x_1'\alpha_1 + \ldots + x_n'\alpha_n.$$

Now

$$\begin{aligned} x_1'\alpha_1 + \ldots + x_n'\alpha_n = \xi\mathbf{A} &= (x_1\alpha_1 + \ldots + x_n\alpha_n)\,\mathbf{A} \\ &= x_1(\alpha_1\mathbf{A}) + \ldots + x_n(\alpha_n\mathbf{A}) \\ &= x_1(\Sigma\alpha_j a_{1j}) + \ldots + x_n(\Sigma\alpha_j a_{nj}) \\ &= (\Sigma x_i a_{i1})\,\alpha_1 + \ldots + (\Sigma x_i a_{in})\,\alpha_n. \end{aligned}$$

Hence, $x_j' = \Sigma x_i a_{ij}$, for $j = 1, 2, \ldots, n$, ... (4)

since $\alpha_1, \ldots, \alpha_n$ form a linearly independent set of vectors in V. The equations (4) can be put in the matrix form

$$X' = XA, \text{ where } A = (a_{ij}). \qquad \ldots (5)$$

Thus the linear operator $\mathbf{A}$ is represented by the matrix A, of order n, with respect to the given basis $\alpha_1, \ldots, \alpha_n$.

Let $\beta_1, \ldots, \beta_n$ be another basis of the linear space V. Then the vectors ξ and $\xi\mathbf{A}$ have different sets of co-ordinate vectors, say $Y = (y_1, \ldots, y_n)$ and $Y' = (y_1', \ldots, y_n')$ respectively, referred to the $[\beta_1, \ldots, \beta_n]$ basis. It then follows from II.2.10 that

$$y_j = \Sigma x_i p_{ij}, \text{ and } y_j' = \Sigma x_i' p_{ij}, \text{ for } j = 1, 2, \ldots n, \qquad \ldots (6)$$

where $(p_{i1}, \ldots, p_{in})$ is the co-ordinate vector of α_i, referred to the $[\beta_1, \ldots, \beta_n]$ basis, i.e., $\alpha_i = p_{i1}\beta_1 + \ldots + p_{in}\beta_n$, for $i = 1, 2, \ldots, n$.

The equations (6) can be put in the matrix form,

$$Y = XP, \text{ and } Y = X'P, \text{ where } P = (p_{ij}). \qquad \ldots (7)$$

As the column vectors of P are the co-ordinate vectors of $\alpha_1, \ldots, \alpha_n$ (referred to the $[\beta_1, \ldots, \beta_n]$ basis); and since $\alpha_1, \ldots, \alpha_n$ are linearly independent in V, it follows that P is non-singular.

It may be noted that the matrix P is uniquely determined when the two bases are given; and conversely, given any non-singular matrix P, and one of the two bases, then the other basis is uniquely determined.

Now, by such a change of basis, the matrix equation (5) for a linear operator A on V is transformed to

$$Y'P=(YP)A, \text{ i.e., } Y'=Y(PAP^{-1}),$$

so that A is represented by the matrix PAP^{-1}, relative to the $[\beta_1,\ldots,\beta_n]$ basis. Thus one obtains the following:

XI.3. *If a linear operator* A *on a linear space V, over a field K, be represented by a matrix A, with respect to a given basis of V, then it is also represented by any matrix similar to A, with respect to a suitable basis of V.*

The set all square matrices of order n, over a field K, is partitioned into disjoint classes of similar matrices. To each linear operator on a linear space V of rank n, over K, there corresponds one such class of similar matrices; and conversely, to each class of similar matrices, there corresponds a linear operator on V.

It can be readily seen that that the null operator O and the identity operator I are represented respectively by the null matrix O and the identity matrix I, with respect to any basis, whatsoever, of V.

Sum and product of linear operators, and multiplicatian of a linear operator by a scalar. Let A, B be two linear operators on V, and let a be any scalar. Then

(i) the operator S, defined by $\alpha\mathrm{S}=\alpha\mathrm{A}+\alpha\mathrm{B}$, for every $\alpha \in V$, is called the *sum of* A *and* B, and is denoted by S$=$A$+$B;

(ii) the operator P, defined by $\alpha\mathrm{P}=(\alpha\mathrm{A})\mathrm{B}$, for every $\alpha \in V$, is called the *product of* A *and* B, and denoted by P$=$AB; and

(iii) the operator C, defined by $\alpha\mathrm{C}=a(\alpha\mathrm{A})$, for every $\alpha \in V$, is called the *product of* A *by the scalar a*, and denoted by C$=$ aA.

It may be readily verified that the operators $\mathbf{S}, \mathbf{P}, \mathbf{C}$, defined above, are linear operators on V. Thus $\mathbf{A}+\mathbf{B}$, $\mathbf{AB}$ and $a\mathbf{A}$ are linear operators. Also the following hold:

(1) $(a+b)\mathbf{A}=a\mathbf{A}+b\mathbf{A}$, (2) $a(\mathbf{A}+\mathbf{B})=a\mathbf{A}+a\mathbf{B}$,

(3) $(ab)\mathbf{A}=a(b\mathbf{A})$ (4) $a(\mathbf{AB})=(a\mathbf{A})\mathbf{B}=\mathbf{A}(a\mathbf{B})$,

(5) $1.\mathbf{A}=\mathbf{A}$, and (6) $0.\mathbf{A}=\mathbf{O}$.

The proof of these results is left as an exercise.

Let $\alpha_1, \ldots, \alpha_n$ be a basis of V. Let the linear operators $\mathbf{A}, \mathbf{B}, \mathbf{S}, \mathbf{P}$ and $\mathbf{C}$ (as defined in (i)-(iii)) be represented by the matrices A, B, S, P and C respectively, and let the co-ordinate vector of α be $X=(x_1, \ldots, x_n)$. Then, it follows directly from the definitions (i), (ii) and (iii) that:

$$XS=XA+XB, \quad XP=(XA)B, \quad \text{and} \quad XC=a(XA),$$

$$\text{i.e.,} \quad XS=X(A+B), \quad XP=X(AB) \quad \text{and} \quad XC=X(aA), \qquad \ldots (8)$$

$$\text{i.e.,} \quad S=A+B, \quad P=AB, \quad \text{and} \quad C=aA,$$

since (8) holds for an arbitrary co-ordinate vector X.

Thus one obtains the following proposition.

X.14 *Relative any basis of a linear space V, over a field K,*

(*i*) *the matrix of the sum of two linear operators is the sum of their matrices;*

(ii) *the matrix of the product of two linear operators is the product of their matrices; and*

(iii) *the matrix of the product of a linear operator by a scalar is the product of its matrix by the same scalar.*

If $\mathbf{A}$ be an invertible linear operator, then its inverse operator $\mathbf{A}^{-1}$ is also a linear operator. Also, from $\mathbf{AA}^{-1}=\mathbf{A}^{-1}\mathbf{A}=\mathbf{I}$, it follows that:

Corollary: The inverse of an invertible linear operator is a linear operator; and their matrices are inverses of one another.

Polynomials in linear operators. Let $\mathbf{A}$ be a linear operator on a linear space V, over a field K, and let $f(x)=a_0+a_1x+a_2x^2+\ldots+a_mx^m$ be a polynomial over K. Then one obtains, correspondingly, a polynomial $f(\mathbf{A})=a_0\mathbf{I}+a_1\mathbf{A}+a_2\mathbf{A}^2+\ldots+a_m\mathbf{A}^m$, in the linear operator $\mathbf{A}$. Referred to a basis of V, the linear operator A is represented by a matrix A, and one obtains the corresponding matric polynomial, $f(A)=a_0I+a_1A+a_2A^2+\ldots+a_mA^m$, in the matrix A.

The characteristic polynomial, the minimal polynomial, the eigen values and the eigen vectors have been defined for matrices. And, since similar matrices have the same characteristic polynomial (vide IX.1.1), the same minimal polynomial (vide IX.1.4); and the same eigen values, these concepts may, as well, be extended to the linear operators. In fact, if A be a linear operator on V, then

(i) $\psi_{\mathrm{A}}(x)=\det(A-xI)$ is defined to be the *characteristic polynomial of the linear operator* A, if A be a matrix, which represents A with respect to a basis of V ;

(ii) if $m(x)$ be a polynomial of the least (positive) degree in $K[x]$, such that $m(\mathrm{A})=O$, then $m(x)$ is called the *minimal polynomial of the linear operator* A ;

(iii) the roots of the characteristic polynomial are called the *eigen values of the linear operator* A ;

(iv) a non-null vector α, such that $\alpha\mathrm{A}=c\alpha$, for some $c\in K$, is called an eigen vector of the linear operator A.

The properties of the characteristic polynomial, the minimal polynomial, eigen values, and eigen vectors for matrices, hold also for linear operators. E.g.

(1) By Cayley-Hamilton theorem $\psi_{\mathrm{A}}(\mathrm{A})=O$;

(2) $f(\mathrm{A})=O$ if, and only if $f(x)$ is divisible by the minimal polynomial $m(x)$ of A.

Exercises

1. Find which of the following are linear operators on the Euclidean 3-space :

(a) $(x, y, z)\,\mathrm{A}=(kx, ky, kz)$, where k is a fixed number.

(b) $(x, y, z)\,\mathrm{B}=(x, 0, 0)$.

(c) $(x, y, z)\,\mathrm{C}=(x, y, 0)$.

(d) $(x, y, z)\mathrm{D}=(x, y, 1)$.

(e) $(x, y, z)\mathrm{E}=(x, y, -z)$.

(f) $(x, y, z)\mathrm{F}=(y, x, z)$.

(g) $(x, y, z)\mathrm{G}=(x+1, y+1, 0)$.

(h) $(x, y, z)\mathrm{H}=(x, (x+y)/2, (x+y+z)/3)$.

(j) $(x, y, z)\,\mathrm{J}=(x \cos \alpha - y \sin \alpha, x \sin \alpha + y \cos \alpha, z)$,

where α has a fixed value.

2. Let V be the linear space generated by the vectors $\alpha=(-1, 2, 0, 1, -1,)$ $\beta=(1, -1, 1, -2, 1)$, and $\gamma=(-1, 1, 0, -1, 0)$.

Let A be the linear operator on V which transforms the vectors α, β, γ respectively to the vectors.

$\alpha_1=(-1, 0, 1, -3, 1)$, $\beta_1=(-2, 3, 0, 0, 1)$, and $\gamma_1=(-1, 1, 1, -1, 0)$.

Find the two matrices A and B, which represent the linear operator A, when referred to the following two bases of V :

(i) α, β, γ, and (ii) $\alpha_2=\beta-\gamma$, $\beta_2=\gamma-\alpha$, $\gamma_2=\alpha+\beta$.

2. Rank and nullity of a linear operator

Let A be a linear operator on a linear space V, over a field K.

Let S be a linear subspace of V. Then its image $S\mathrm{A}$ and the pre-image $S\mathrm{A}^{-1}$ are both linear subspaces of V.

In fact, if $x\mathrm{A}, y\mathrm{A} \in S\mathrm{A}$ i.e., $x, y \in S$, and if a be any scalar then $x\mathrm{A}+y\mathrm{A}=(x+y)\mathrm{A} \in S\mathrm{A}$, and $a(x\mathrm{A})=(ax)\mathrm{A} \in S\mathrm{A}$ (since $x+y$ and ax belong to S, as S is a linear subspace). Thus $S\mathrm{A}$ is a linear subspace of V. Similarly, $S\mathrm{A}^{-1}$ is also a linear subspace of V.

Thus one obtains the following :

X.2.1. *The image and the pre-image of a linear subspace, under any linear operator, are also linear subspaces.*

The image of the entire linear space V, under a linear operator A, is called the *range of the linear operator* A. It is a subspace of V ; and its rank is called the *rank of the linear operator* A.

The pre-image of the null vector θ, under a linear operator A, is called the *kernel* or the *null space of the linear operator* A. It is a subspace of V, and its rank is called the *nullity of the linear operator* A. ([illegible])

X.2.2. *The sum of the rank and nullity of a linear operator* A *on a linear space V, over a field K, is equal to the rank of V.*

Proof: Let N and R denote respectively the null space and the range of A on V: and let rank $(N)=s$, and rank $(R)=r$.

Let $\alpha_1, \dots, \alpha_r$ form a basis of R, and let $\beta_1, \dots, \beta_r$ be the pre-images of $\alpha_1, \dots, \alpha_r$ respectively, under A. Then the vectors $\beta_1, \dots, \beta_r$ are linearly independent in V, since

$$b_1\beta_1+\dots+b_r\beta_r=\theta \text{ implies } (b_1\beta_1+\dots+b_r\beta_r)\mathrm{A}=\theta,$$

i.e., $$b_1\alpha_1+\dots+b_r\alpha_r=\theta, \text{ so that } b_1=\dots=b_r=0,$$

since $\alpha_1, \dots, \alpha_r$ are linearly independent.

Let S be the subspace of V, generated by the vectors $\beta_1, \dots, \beta_r$. Then $\beta_1, \dots, \beta_r$ form a basis of S, as these vectors are linearly independent; hence rank $(S)=r$.

It will now be shown that V is the direct sum of the subspaces S and N. Firstly, let $\gamma \in S \cap N$. Then γ is expressible as $\gamma=c_1\beta_1+\dots+c_r\beta_r$; and also $\gamma\mathrm{A}=\theta$. Thus

$$(c_1\beta_1+\dots+c_r\beta_r)\mathrm{A}=\theta, \text{ i.e., } c_1\alpha_1+\dots+c_r\alpha_r=\theta,$$

so that $c_1=\dots=c_r=0$ (since $\alpha_1, \dots, \alpha_r$ are linearly independent). Consequently, $\gamma=\theta$; thus $S \cap N=(\theta)$.

Next, let λ be an arbitary vector in V. Then $\lambda\mathrm{A} \in R$, and so $\lambda\mathrm{A}$ is expressible as

$$\lambda\mathrm{A}=g_1\alpha_1+\dots+g_r\alpha_r, \text{ where } g_1, \dots, g_r \in K.$$

Let μ be the vector in S, defined by

$$\mu=g_1\beta_1+\dots+g_r\beta_r,$$

then $$\mu\mathrm{A}=(g_1\beta_1+\dots+g_r\beta_r)\mathrm{A}=g_1\alpha_1+\dots+g_r\alpha_r=\lambda\mathrm{A}.$$

Hence $(\lambda-\mu)\mathrm{A}=\lambda\mathrm{A}-\mu\mathrm{A}=\theta$, so that $\sigma=\lambda-\mu \in N$.

Thus $\lambda=\mu+\sigma$, where $\mu \in S$ and $\sigma \in N$.

V is therefore the direct sum of the subspaces S and N. Hence

$$\operatorname{rank}(V)=\operatorname{rank}(S)+\operatorname{rank}(N)=r+s=\operatorname{rank}(\mathrm{A})+\operatorname{nullity}(\mathrm{A}).$$

X.2.3. *The rank of a linear operator* A *on a linear space* V, *over a field* K, *is equal to the rank of the matrix that represents* A, *with respect to any basis of* V.

Proof : Let $\alpha_1, \ldots, \alpha_n$ be a basis of V. Then every vector of V is expressible as $a_1\alpha_1 + \ldots + a_n\alpha_n$, where the coefficients a_i are taken from K.

Let R denote the range of A. Then R consists of all vectors,

$$(a_1\alpha_1 + \ldots + a_n\alpha_n)\mathrm{A} = a_1(\alpha_1\mathrm{A}) + \ldots + a_n(\alpha_n\mathrm{A}),$$

where a_i run over K. Then R is the subspace of V generated by the vectors $\alpha_1\mathrm{A}, \ldots, \alpha_n\mathrm{A}$. Hence the rank of R, i.e, the rank of the linear operator A, is given by the number of linearly independent vectors in the set $\alpha_1\mathrm{A}, \ldots, \alpha_n\mathrm{A}$.

Again, if A be the matrix that represents the linear operator A, referred to the basis $\alpha_1, \ldots, \alpha_n$, then the successive row-vectors of A are given by the co-ordinate vectors of the vectors $\alpha_1\mathrm{A}, \ldots, \alpha_n\mathrm{A}$ (referred to the basis $\alpha_1, \ldots, \alpha_n$) (vide proof of X.1.2). Also, the number of linearly independent row-vectors of A gives the rank of the matrix A.

Hence the rank of the linear operator A is equal to the rank of the matrix A. And, the value of the rank is independent of the particular basis that might have been used ; that is, it is the same for every basis of V.

For a square matrix A, of order n, the nullity has been defined by : nullity of $A = n -$ rank of A. Hence, it now follows from X.2.2 and X.2.3 that *the nullity of a linear operator* A *on a linear space* V, *over a field* K, *is equal to the nullity of any matrix that represents* A, *referred to some basis of* V.

An invertible linear operator is also known as a *non-singular linear operator*. Let A be a non-singular linear operator on a linear space V, over a field K. Then A is a bijective mapping of V onto the whole of V. Hence the range of A is the entire linear space V, and the null space of A consists of the null vector θ only. The nullity of A is 0, and the rank of A, and hence also the rank of any matrix representing A (with respect to some basis of V) is n. Consequently, any matrix that represents a non-singular linear operator is a non-singular matrix. Each of the above properties completely characterises a non-singular linear operator.

Isomorphism of a linear space V_1 onto a linear space V_2, both having the same field of scalars, has been defined in page 37. A non-singular linear operator is found to be an isomorphism of the linear space *onto the whole of itself*. Such an isomorphism is called an *automorphism of the linear space*. Thus every linear operator on a linear space V, over a field K, is an automorphism of V.

Exercises

1. Find the rank and nullity of the linear operators defined in exercise 1, pages 236-7.

2. Given two subspaces N and R of a linear space V of rank n, over a field K, there exists a linear operator $\mathbf{A}$ on V, such that N and R are respectively the null space and the range of $\mathbf{A}$.

[*Hints* : Let $\text{rank}(N)=s$, and $\text{rank}(R)=r$. Let $\alpha_1,\ldots,\alpha_r$ form a basis of R, and $\beta_1,\ldots,\beta_s$ form a basis of N. Construct a basis $\beta_1,\ldots,\beta_n$ of V, by extending the basis of N. By X.1.1, there exists a linear operator $\mathbf{A}$ on V, defined by

$$\beta_i\mathbf{A}=\theta, \text{ for } i=1,\ldots s, \text{ and } \beta_i\mathbf{A}=\alpha_i, \text{ for } i=s+1,\ldots n.$$

Then $\mathbf{A}$ is the required linear operator.]

3. Invariant subspaces

Let $\mathbf{A}$ be a linear operator on a linear space V, over a field K. A linear subspace S of V is said to be an *invariant subspace* of $\mathbf{A}$, if $S\mathbf{A} \subset S$.

The entire linear space V, and the null space of $\mathbf{A}$ (which may even consist of the null vector θ only) are evidently invariant subspaces of $\mathbf{A}$

Invariant subspaces of rank 1 are of particular interest. Let S be an invariant subspace of rank 1. Then S consists of all vectors $r\alpha$, where α is a non-null vector in V, and r runs over K. Also, from $SA \subset S$, it follows that $\alpha A = \lambda\alpha$, for some $\lambda \varepsilon K$. α is therefore an eigen vector of A, having λ as the corresponding eigen value. The converse is also true. That is, if α be an eigen vector of A, then the linear subsace of rank 1, generated by α, is an invariant subspace of A.

Let λ be an eigen value of A, and let S be the characteristic subspace, corresponding to λ. Every vector, belonging to S, is an eigen vector of A (corresponding to λ). Hence, S is an invariant subspace of A; and every linear subspace of rank 1 of S is also an invariant subspace of A.

X.3.1. *If S be an invariant subspace of* A, *then it is also an invariant subspace of* $f(A)$, *where* $f(x) \varepsilon K[x]$.

Proof: Let $f(x) = a_0 + a_1x + a_2x^2 + \ldots + a_mx^m \varepsilon K[x]$. Then $f(A) = a_0I + a_1A + a_2A^2 + \ldots a_mA^m$. By hypothesis, $SA \subset S$, from which it follows successively that.

$$SA^2 = (SA)A \subset SA \subset S, \quad SA^3 = (SA^2)A \subset SA \subset S, \ldots, SA^m \subset S.$$

$$\begin{aligned} \text{Hence} \quad Sf(A) &= S(a_0I + a_1A + a_2A^2 + \ldots + a_mA^m) \\ &= a_0S + a_1SA + a_2SA^2 + \ldots + a_mSA^m \\ &\subset a_0S + a_1S + a_2S + \ldots + a_mS \\ &= (a_0 + a_1 + \ldots + a_m)S = aS \subset S. \end{aligned}$$

Of particular importance is the following proposition, which gives a convenient method for finding out invariant subspaces of a given linear operator.

X.3.2. *For any polynomial* $f(x) \varepsilon K[x]$, *the null space of* $f(A)$ *is an invariant subspace of the linear operator* A.

Proof: Let N be the null space of $f(A)$, so that $Nf(A) = \theta$. Also, the linear operator A is permutable with $f(A)$, i.e., $A.f(A) = f(A).A$. Now,

$$(NA)f(A) = N(A.f(A)) = N(f(A).A) = (N.f(A))A = \theta A = \theta,$$

from which, it follows that $NA \subset N$; N is therefore an invariant subspace of A.

Let S be an invariant subspace of A. Let A′ be a linear operator on S, defined by $\alpha A' = \alpha A$, for every $\alpha \varepsilon S$. Then A′ is called the *induced* linear operator on S.

A linear operator **A** is represented by a matrix A, when referred to a particular basis of the underlying linear space V. It is also represented by every matrix PAP^{-1} (where P is non-singular), that is by every matrix similar to A, as one refers to different bases of V. Hence the problem of finding out the *simplest* matrix, representing the given linear operator **A**, is identical with the problem of finding out the normal form of the matrix A (which represents **A** with respect some basis of V) under similarity. The latter problem has been solved in Art. 4, Chapter IX, by the methods of matrix calculus. The same can as well be done, by utilising the invariant subspaces of **A**; the method brings out the geometrical aspects, and therefore appeals more to intuition. A brief outline of the method is given below.

X.3.3. *If S be an invariant subspace of a linear operator* **A** *on a linear space V over a field K, then referred to a basis $\alpha_1,\ldots,\alpha_r,\ldots,\alpha_n$ of V, where $\alpha_1,\ldots,\alpha_r$ form a basis of S,* **A** *is represented by a matrix A of the form*

$$A=\begin{pmatrix} B & O \\ G & C \end{pmatrix}, \qquad \ldots \quad (1)$$

where B and C are square matrices, of which B is of order r, and and G and O are rectangular matrices of suitable sizes, O being a null matrix.

Proof: To obtain the matrix which represents **A**, with respect to basis $\alpha_1,\ldots,\alpha_n$, one has to express the vectors $\alpha_1\mathbf{A},\ldots,\alpha_n\mathbf{A}$ as linear combinations of the basis vectors. Then the coefficients of these linear combinations form the required matrix.

Since S is an invariant subspace of **A**, $S\mathbf{A}\subset S$. Consequently, when $\alpha_1\mathbf{A},\ldots,\alpha_r\mathbf{A}$ are expressed as linear combinations of the basis vectors $\alpha_1,\ldots,\alpha_r,\ldots,\alpha_n$, the coefficients for $\alpha_{r+1},\ldots,\alpha_n$ are all null. Hence

$$\begin{aligned}\alpha_1\mathbf{A}&=b_{11}\alpha_1+\ldots+b_{1r}\alpha_r\\ \ldots\quad&\ldots\quad\ldots\quad\ldots\\ \alpha_r\mathbf{A}&=b_{r1}\alpha_1+\ldots+b_{rr}\alpha_r,\end{aligned}$$

where b_{ij} are suitable scalars. Also, let

$$\begin{aligned}\alpha_{r+1}\mathrm{A} &= g_{11}\alpha_1+\ldots+g_{1r}\alpha_r+c_{11}\alpha_{r+1}+\ldots+c_{1s}\alpha_n\\ \ldots\quad & \ldots\quad \ldots\quad \ldots\quad \ldots\\ \alpha_n\mathrm{A} &= g_{s1}\alpha_1+\ldots+g_{sr}\alpha_r+c_{s1}\alpha_{r+1}+\ldots+c_{ss}\alpha_n,\end{aligned}$$

where g_{ij} and c_{ij} are suitable scalars, and $s=n-r$.

The linear operator A is then represented by the matrix A, of the form (1), where $B=(b_{ij})$ is an $r\times r$ matrix, $C=(c_{ij})$ is an $s\times s$ matrix, $G=(g_{ij})$ is an $s\times r$ matrix, and O is an $r\times s$ null matrix.

A matrix A of the form (1) is said to be *semidecomposable*.

The converse of the above proposition is also true. In fact,

X.3.4. *If a linear operator* A *on a linear space V, over a field K, be represented by a semidecomposable matrix A, as given in* (1), *with respect to a basis $\alpha_1,\ldots,\alpha_n$, then the linear subspace of V, generated by the first r vectors $\alpha_1,\ldots,\alpha_r$, is an invariant sub-space of* A.

Proof: It follows from the form of the matrix A, as given in (1), representing the linear operator A, that

$$\begin{aligned}\alpha_1\mathrm{A} &= b_{11}\alpha_1+\ldots+b_{1r}\alpha_r\\ \ldots\quad & \ldots\quad \ldots\quad \ldots\\ \alpha_r\mathrm{A} &= b_{r1}\alpha_1+\ldots+b_{rr}\alpha_r.\end{aligned}$$

Again, every vector in S can be expressed as a linear combination, $h_1\alpha_1+\ldots+h_r\alpha_r$; and

$$\begin{aligned}(h_1\alpha_1+\ldots+h_r\alpha_r)\mathrm{A} &= h_1(\alpha_1\mathrm{A})+\ldots+h_r(\alpha_r\mathrm{A})\\ &= h_1(b_{11}\alpha_1+\ldots+b_{1r}\alpha_r)+\ldots+h_r(b_{r1}\alpha_1+\ldots+b_{rr}\alpha_r)\\ &= (h_1b_{11}+\ldots+h_rb_{r1})\alpha_1+\ldots+(h_1b_{1r}+\ldots+h_rb_{rr})\alpha_r \in S.\end{aligned}$$

Hence S is an invariant subspace of A.

The results of X.3.3. and X.3.4 can be sharpened to the following form :

X.3.5. *Let* A *be a linear operator on a linear space V of rank n, over a field K.*

(i) *If V be the direct sum of the linear subspaces $S_1,\ldots,S_m$ of ranks $r_1,\ldots,r_m$ respectively, where each S_i is an invariant sub-space of* A, *then referred to a basis,*

$$\alpha_1,\ldots,\alpha_n,\ (n=r_1+\ldots+r_m), \qquad \ldots (2)$$

such that the first r_1 vectors in (2) *form a basis of* S_1, *the next* r_2 *vectors in* (2) *form a basis of* S_2, *and so on, and finally the last* r_m *vectors in* (2) *form a basis of* S_m, *the linear operator* A *is represented by a block-diagonal matrix,*

$$A = \text{diag}\,(A_1, \ldots, A_m), \text{ where } A_i \text{ is of order } r_i, \quad \ldots (3)$$

for $i = 1, \ldots, m$.

(ii) *Conversely, if the linear operator* A *be represented by a block-diagonal matrix* A, *given in* (3), *with respect to a basis* (2), *and if* S_1 *be the linear subspace generated by the first* r_1 *vectors in* (2), S_2 *be the linear subspace generated by the next* r_2 *vectors in* (2), *and so on, and finally* S_m *be the linear subspace generated by the last* r_m *vectors in* (2), *then each of* $S_1, \ldots, S_m$ *is an invariant subspace of the linear operator* A, *and* V *is the direct sum of* $S_1, \ldots, S_m$.

Proof: (i) Since S_1 is an invariant subspace of A, the vectors $\alpha_1 A, \ldots, \alpha_{r_1} A$ lie in S_1, and are expressible uniquely as linear combinations of the vectors $\alpha_1, \ldots \alpha_{r_1}$. Thus when the vectors $\alpha_1 A, \ldots, \alpha_{r_1} A$ are expressed as linear combinations of the basis vectors (2), the coefficients of $\alpha_{r_1+1}, \ldots, \alpha_n$ in each such linear combination shall be null ; thus

$$\alpha_1 A = a_{11}\alpha_1 + \ldots + a_{1r_1}\alpha_{r_1}$$

$$\ldots \quad \ldots \quad \ldots \quad \ldots$$

$$\alpha_{r_1} A = a_{r_1 1}\alpha_1 + \ldots + a_{r_1 r_1}\alpha_{r_1}.$$

Similarly, since S_2 is an invariant subspace of A, when the vectors $\alpha_{r_1+1} A, \ldots, \alpha_{r_1+r_2} A$ are expressed as linear combinations of the basis vectors (2), the coefficients of $\alpha_1, \ldots, \alpha_{r_1}$, $\alpha_{r_1+r_2+1}, \ldots, \alpha_n$ shall be all null Thus

$$\alpha_{r_1+1} A = a_{r_1+1, r_1+1}\alpha_{r_1+1} + \ldots + a_{r_1+1, r_1+r_2}\alpha_{r_1+r_2}$$

$$\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$

$$\alpha_{r_1+r_2} A = a_{r_1+r_2, r_1+1}\alpha_{r_1+1} + \ldots + a_{r_1+r_2, r_1+r_2}\alpha_{r_1+r_2}.$$

And, so on, and finally since S_m is an invariant subspace of A, when the vectors $\alpha_{t+1} A, \ldots, \alpha_n A$ (where $t = r_1 + \ldots + r_{m-1}$) are expressed as linear combinations of the basis vectors (2), the coefficients of $\alpha_1,\ldots,\alpha_t$ shall be all null. Thus

$$\alpha_{t+1}\mathbf{A}=a_{t+1\,t+1}\alpha_{t+1}+\ldots+a_{t+1\,n}\alpha_n.$$

$$\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$

$$\alpha_n\mathbf{A}\ =a_{n\,t+1}\alpha_{t+1}+\ldots+a_{nn}\alpha_n.$$

All the coefficients a_{ij} occuring above are suitable scalars from the ground field K.

It follows from above that the linear operator $\mathbf{A}$ is represented by the block-diagonal matrix A, given in (3), when referred to a basis (2) as described above.

(ii) It can be proved, exactly in the same way as in the case of X.3.4, that $S_1,\ldots,S_m$ are invariant subspaces of $\mathbf{A}$. Also, any vector α in V can be expressed uniquely as a linear combination of the basis vectors (2),

$$\begin{aligned}\alpha&=a_1\alpha_1+\ldots+a_n\alpha_n\\&=(a_1\alpha_1+\ldots+a_{r_1}\alpha_{r_1})+\ldots+(a_{t+1}\alpha_{t+1}+\ldots+a_n\alpha_n)\\&=\beta_1+\ldots+\beta_m,\end{aligned}$$

where $\beta_1 \in S_1,\ldots, \beta_m \in S_m$. Hence V is a linear sum of $S_1,\ldots,S_m$. And, since the above representation of α as a sum of the vectors $\beta_1,\ldots,\beta_m$ belonging to the linear subspaces $S_1,\ldots,S_m$ is unique, it follows that V is the direct sum of $S_1,\ldots,S_m$.

Let $X=(x_1,\ldots,x_n)$ be the co-ordinate vector representing an arbitrary vector $\xi \in V$. If A be the matrix in the block-diagonal form (3), then in the matrix product,

$$XA=(x_1,\ldots,x_n).\mathrm{diag}(A_1,\ldots,A_m),$$

the effect of A on the first r_1 co-ordinates is the same as that of the block A_1, the effect of A on the next r_2 co-ordinates is the same that of the block A_2, and so on, and finally the effect of A on the last r_m co-ordinates is given by that of the block A_m. Hence it follows that *the diagonal blocks* $A_1,\ldots,A_m$ *represent the induced linear operators* $\mathbf{A}_1,\ldots,\mathbf{A}_m$ *on the invariant subspaces* $S_1,\ldots,S_m$, *referred to the bases* $[\alpha_1,\ldots,\alpha_{r_1}],\ldots,$ $[\alpha_{t+1},\ldots,\alpha_n]$ *respectively.*

It follows from the above proposition that to obtain a matrix in the block-diagonal form (3), representing a given linear operator **A**, it is necessary to find out some invariant subspaces of **A** whose direct sum is the entire linear space V. It will now be shown how such invariant subspaces may be found out.

It can be shown that *if* $f(x)$ *and* $g(x)$ *be two polynomials, relatively prime in* $K[x]$, *then the null space of* $f(\mathbf{A}).g(\mathbf{A})$ *is the direct sum of the null spaces of* $f(\mathbf{A})$ *and* $g(\mathbf{A})$. (This can be proved in several steps, as indicated in exercise 2 below).

Let $m(x)$ be the minimal polynomial of **A**. Then $m(\mathbf{A})=\mathbf{O}$, so that $V.m(\mathbf{A})=V\mathbf{O}=\theta$; thus V is the null space of $m(\mathbf{A})$. Let $m(x)=g_1(x)\ldots g_m(x)$, where the polynomlals $g_1(x),\ldots,g_m(x)$ are prime in pairs in $K[x]$. Let $S_1,\ldots,S_m$ be the null spaces of $g_1(\mathbf{A}),\ldots,g_m(\mathbf{A})$ respectively. Then, by the result stated above, V is the direct sum of the linear subspaces $S_1,\ldots,S_m$. Also, by X.3.2, each of $S_1,\ldots,S_m$ is an invariant subspace of **A**. Thus V is the direct sum of the linear subspaces $S_1,\ldots,S_m$ which are invariant subspaces of **A**. Then **A** is represented by a block-diagonal matrix A, as in (3), when referred to a basis (2), which is formed by collecting together a basis of S_1, then a basis of S_2, and so on, and finally a basis of S_m.

The diagonal blocks $A_1,\ldots,A_m$ are the matrices representing the induced linear operators $\mathbf{A}_1,\ldots,\mathbf{A}_m$ on the invariant subspaces $S_1,\ldots S_m$, referred to their respective bases as selected above. Next, one considers each invariant subspace S_i and the induced linear operator $\mathbf{A}_i$ on it separately, for $i=1,\ldots m$. To form the basis (2) for V, the bases for $S_1,\ldots,S_m$ were chosen in an arbitrary manner. Now, for each S_i, it is generally possible to choose a suitable basis, referred to which the induced linear operator $\mathbf{A}_i$ will be represented by a matrix, much simpler in form than the block A_i in (3). And by doing so, for each of the invariant subspaces $S_1,\ldots,S_m$ and collecting together these bases, one obtains a basis of V, referred to which, the linear operator **A** is represented by a block-diagonal matrix, whose blocks are much simpler in nature than those in (3).

Normal forms of matrices, under similarity, may thus be obtained by choosing suitable bases as indicated above. Since these normal forms have already been obtained in a different manner (by using the methods of matrix calculus), the method outlined above is not being elaborated here. Intending readers are referred to standard books on Linear Algebra, where this method has been used to obtain the normal forms of matrices.

EXAMPLE 1: Let **A** be a linear operator on the unitary space $V_4(E)$ represented by the matrix

$$\begin{pmatrix} 1 & 0 & 2 & -3 \\ 0 & 1 & 4 & -6 \\ 3 & -1 & 0 & 2 \\ 3 & -1 & -1 & 3 \end{pmatrix}.$$

when referrrd to the basis $\alpha_1, \alpha_2, \alpha_3, \alpha_4$. Show that the subspace S, spanned by the vectors $2\alpha_1 - \alpha_2$ and $\alpha_3 - \alpha_4$, is an invariant subspace of **A**.

Soln: It follows from the the matrix representation of **A** that

$$\alpha_1 A = \alpha_1 + 2\alpha_3 - 3\alpha_4, \qquad \alpha_2 A = \alpha_2 + 4\alpha_3 - 6\alpha_4,$$

$$\alpha_3 A = 3\alpha_1 - \alpha_2 + 2\alpha_4, \qquad \alpha_4 A = 3\alpha_1 - \alpha_2 - \alpha_3 + 3\alpha_4.$$

Any vector in **S** is of the form $a(2\alpha_1 - \alpha_2) + b(\alpha_3 - \alpha_4)$; and

$$\begin{aligned}[a(2\alpha_1 - \alpha_2) + b(\alpha_3 - \alpha_4)]\,A &= 2a(\alpha_1 A) - a(\alpha_2 A) + b(\alpha_3 A) - b(\alpha_4 A) \\ &= 2a(\alpha_1 + 2\alpha_3 - 3\alpha_4) - a(\alpha_2 + 4\alpha_3 - 6\alpha_4) \\ &\qquad + b(3\alpha_1 - \alpha_2 + 2\alpha_4) - b(3\alpha_1 - \alpha_2 - \alpha_3 + 3\alpha_4) \\ &= a(2\alpha_1 - \alpha_2) + b(\alpha_3 - \alpha_4) \;\varepsilon\; S.\end{aligned}$$

Hence S is an invariant subspace of **A**.

EXAMPLE 2: *If* **A** *be a linear operator on a linear space V, over the complex field C (or any algebraically closed field), then every invariant subspace S of* **A** *contains an eigen vector of* **A**.

Soln: Let **B** be the induced linear operator on S. Then, referred to a basis $\alpha_1, \ldots, \alpha_r$ of S, **B** is represented by a matrix B of order r, over C. The roots of the characteristic polynomial of **B**, i.e., the eigen values of **B** lie in C (as C is algebraically closed). Let λ be an eigen value, then the co-ordinate vectors of the eigen vectors of **B**, corresponding to λ, have complex co-ordinates (since these are given by the non-null solutions of a system of linear equations with complex coefficients). If $c_1, \ldots, c_r$ be the co-ordinates of such an eigen vector β, then $\beta = c_1\alpha_1 + \ldots + c_r\alpha_r \;\varepsilon\; S$.

EXAMPLE 3: *If* **A** *and* **B** *be two linear operators on a linear space V, over a field K, such that* **AB**=**BA**, *then*

(i) *the characteristic subspace of an eigen value λ of* **A** *is an invariant subspace of* **B**;

(ii) *if, moreover the ground field K be the complex field C (or any algebraically closed field), then* **A** *and* **B** *have a common eigen vector.*

Soln : (i) Let S be the characteristic subspace, corresponding to the eigen value λ of A; and let $\alpha \varepsilon S$ then $\alpha A = \lambda\alpha$. Now,

$$(\alpha B)A = \alpha(BA) = \alpha(AB) = (\alpha A)B = (\lambda\alpha)B = \lambda(\alpha B),$$

so that $\alpha B \varepsilon S$. Hence S is invariant subspace of A.

(ii) As S is an invariant subspace of B, it follows by example 2, that S contains an eigen vector β of B. Again, since S is a characteristic subspace, corresponding to an eigen value λ of A, every vector belonging to S is an eigen vector of A. Hence β is a common eigen vector of A and B.

Exercises

1. Let A be a linear operator on a linear space V, over a field K and let S be an invariant subspace of A. Then

(i) the image and pre-image of S, under A, are also invariant subspaces of A ;

(ii) S is also an invariant subspace of A^{-1}, if A is non-singular.

2. Let A be a linear operator on a linear space V, over a field K. Let $f(x)$ and $g(x)$ be two polynomials over K. Show that :

(i) if $h(x)$ be the h.c.f. of $f(x)$ and $g(x)$, then the null space of $h(A)$ is the intersection of the null spaces of $f(A)$ and $g(A)$;

(ii) if $c(x)$ be the l.c.m. of $f(x)$ and $g(x)$, then the null space of $c(A)$ is the linear sum of the null spaces of $f(A)$ and $g(A)$;

(iii) if $f(x)$ and $g(x)$ be relatively prime, then the null space of $f(A).\, g(A)$ is the direct sum of the null spaces of $f(A)$ and $g(A)$.

3. Let A be a matrix in the Jordan normal form, repesenting a linear operator A on a linear space V, over a field K, referred to a particular basis.

Find the induced linear operators, represented by the different Jordan blocks, on the invariant subspaces distinguished by the different blocks. Determine the characteristic properties of the particular basis vectors for each of these invariant subspaces.

4. Adjoint of a linear operator.

Let V be a unitary space of rank n (over the complex field C). The scalar product of two vectors α, $\beta \varepsilon V$ will henceforth be denoted by $[\alpha, \beta]$, instead of $\alpha.\beta$ as was being done so long.

Let A be a linear operator on V. Then an operator B on V, defined by

$$[\alpha \mathbf{A}, \beta] = [\alpha, \beta \mathbf{B}], \text{ for every pair of vectors } \alpha, \beta \varepsilon V, \quad \text{---}(1)$$

is called the *adjoint of* A, and will be denoted by $\mathbf{B} = \mathbf{A}^\circ$.

The adjoint operator B is necessarily linear. In fact, if α, β, $\gamma \epsilon V$, and if b be an arbitrary scalar, then it follows, by II.3.6 (i)-(ii), that

$$[\alpha, (\beta+\gamma)\mathbf{B}] = [\alpha\mathbf{A}, \beta+\gamma] = [\alpha\mathbf{A}, \beta] + [\alpha\mathbf{A}, \gamma] = [\alpha, \beta\mathbf{B}] + [\alpha, \gamma\mathbf{B}],$$

$$[\alpha, (b\beta)\,\mathbf{B}] = [\alpha\mathbf{A}, b\beta] = b^*[\alpha\mathbf{A}, \beta] = b^*\,[\alpha, \beta\mathbf{B}] = [\alpha, b(\beta\mathbf{B})].$$

X.4.1. *If* A, B *be any two linear operators on a unitary space* V, *and if* b *be any scalar, then*

(*i*) $(\mathbf{A}^\circ)^\circ = \mathbf{A}$, (*ii*) $(b\mathbf{A})^\circ = b^*\mathbf{A}^\circ$, (*iii*) $\mathbf{I}^\circ = \mathbf{I}$,

(*iv*) $(\mathbf{A}+\mathbf{B})^\circ = \mathbf{A}^\circ + \mathbf{B}^\circ$, *and* (*v*) $(\mathbf{AB})^\circ = \mathbf{B}^\circ\mathbf{A}^\circ$.

Proof: (*i*) From $[\alpha\mathbf{A}, \beta] = [\alpha, \beta\mathbf{A}^\circ]$, one obtains $[\beta, \alpha\mathbf{A}] = [\beta\mathbf{A}^\circ, \alpha]$. Interchanging α and β, one gets $[\alpha, \beta\mathbf{A}] = [\alpha\mathbf{A}^\circ, \beta]$; that is, $[\alpha\mathbf{A}^\circ, \beta] = [\alpha, \beta\mathbf{A}]$, from which it follows that $(\mathbf{A}^\circ)^\circ = \mathbf{A}$.

(*ii*) $$[\alpha, \beta(b\mathbf{A})^\circ] = [\alpha(b\mathbf{A}), \beta] = b[\alpha\mathbf{A}, \beta] = b[\alpha, \beta\mathbf{A}^\circ]$$
$$= [\alpha, b^*\beta\mathbf{A}^\circ] = [\alpha, \beta(b^*\mathbf{A})].$$

Thus $(b\mathbf{A})^\circ = b^*\mathbf{A}^\circ$.

(*iii*) $[\alpha, \beta\mathbf{I}^\circ] = [\alpha\mathbf{I}, \beta] = [\alpha, \beta]$; thus $\mathbf{I}^\circ = \mathbf{I}$.

(*iv*) $$[\alpha, \beta(\mathbf{A}+\mathbf{B})^\circ] = [\alpha(\mathbf{A}+\mathbf{B}), \beta] = [\alpha\mathbf{A}, \beta] + [\alpha\mathbf{B}, \beta]$$
$$= [\alpha, \beta\mathbf{A}^\circ] + [\alpha, \beta\mathbf{B}^\circ] = [\alpha, \beta(\mathbf{A}^\circ + \mathbf{B}^\circ)].$$

Thus $(\mathbf{A}+\mathbf{B})^\circ = \mathbf{A}^\circ + \mathbf{B}^\circ$.

(*v*) $$[\alpha, \beta(\mathbf{AB})^\circ] = [\alpha(\mathbf{AB}), \beta] = [(\alpha\mathbf{A})\mathbf{B}, \beta] = [\alpha\mathbf{A}, \beta\mathbf{B}^\circ]$$
$$= [\alpha, (\beta\mathbf{B}^\circ)\mathbf{A}^\circ] = [\alpha, \beta(\mathbf{B}^\circ\mathbf{A}^\circ)].$$

Thus $(\mathbf{AB})^\circ = \mathbf{B}^\circ\mathbf{A}^\circ$.

The existence and uniqueness of the adjoint of a given linear operator are established in the following proposition :

X.4.2. *If* **A** *be a linear operator on a unitary space* V *of rank* n, *then a linear operator* **B** *on* V *is the adjoint of* **A** *if, and only if, the matrix representing* **B** *is the conjugate transpose of the matrix representing* **A**, *when referred to an orthonormal basis of* V.

Proof: Let $\alpha_1,\ldots,\alpha_n$ form an orthonormal basis of V, and let **A** be represented by the matrix $A=(a_{ij})$, when referred to this basis. Let ξ, η be two vectors in V with the respective co-ordinate vectors $X=(x_1,\ldots,x_n)$ and $Y=(y_1,\ldots,y_n)$, with respect to the above basis. Then $[\xi\mathbf{A},\eta]$ is represented by

$$[XA, Y]=[Z, Y]=z_1y_1^*+\ldots+z_ny_n^*,$$

where $(z_1,\ldots,z)=Z=XA=(x_1,\ldots,x_n)\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}$

i e., $z_j=x_1a_{1j}+x_2a_{2j}+\ldots+x_na_{nj}$, for $j=1,\ldots,n$.

Thus $[\xi\mathbf{A},\eta]$ is represented by the bilinear form

$$\Sigma(\Sigma x_ia_{ij})y_j^*=\Sigma x_ia_{ij}y_j^*=XA^tY^{*t}, \qquad \ldots \qquad (2)$$

whose matrix is the transpose of the matrix A (representing **A**).

Again, the bilinear form (2) can be expressed as

$$XA^tY^{*t}=\Sigma x_ia_{ij}y_j^*=\Sigma x_i(\Sigma a_{ij}y_j^*)$$

$$=x_1w_1^*+x_2w_2^*+\ldots+x_nw_n^*=[X, W]$$

where $w_i^*=\Sigma a_{ij}y_j^*$, i.e , $w_i=\Sigma a_{ij}^*y_j=\Sigma y_ja_{ij}^*$,

i e., $W=(w_1,\ldots,w_n)=(y_1,\ldots,y_n)\begin{pmatrix} a_{11}^* & a_{21}^* & \ldots & a_{n1}^* \\ a_{12}^* & a_{22}^* & \ldots & a_{n2}^* \\ \ldots & & \ldots & \ldots \\ a_{1n}^* & a_{2n}^* & \ldots & a_{nn}^* \end{pmatrix}$

$=YA^{*t}$.

Thus $XA^tY^{*t}=[X, YA^{*t}]=[X, YB]$, if, and only if, $B=A^{*t}$.

Hence the bilinear form (2) also represents $[\xi, \eta\mathbf{B}]$ if, and only if, the linear operator **B** is represented by the matrix A^{*t}.

Thus $[\xi\mathbf{A}, \eta]=[\xi,\eta\mathbf{B}]$ holds for every pair of vectors $\xi, \eta\ \varepsilon\ V$; that is, **B** is the adjoint of **A** if, and only if, the matrix represening **B** is the conjugate transpose of the matrix representing **A**, with respect to an orthonormal basis of V.

Corollary: *There exists a one-to-one correspondence between bilinear forms and linear operators on a unitary space V, such that, when referred to an orthonormal basis of V, corresponding to a linear operator* **A**, *represented by the matrix A, there exists a bilinear form XA^tY^{*t}, satisfying the relations,*

$$XA^tY^{*t}=[XA, Y]=[X, YA^{*t}].$$

where X, Y are the co-ordinate vectors representing a pair of arbitrary vectors $\xi, \eta \in V$.

Exercises

1. A linear operator **A**, on a unitary space V of rank 3, is represented by the following matrix, with respect to an orthonormal basis α, β, γ,

$$\begin{pmatrix} 2 & 1-i & 3i \\ 1+i & 2i & 0 \\ -i & 5 & 2+i \end{pmatrix}.$$

Find the matrices representing **A** and $\mathbf{A}^\circ$, with respect to the basis $\alpha+\beta, \beta-\gamma, \gamma$; and examine whether the matrix for $\mathbf{A}^\circ$ is the conjugate transpose of the matrix for **A**.

5. Hermitian and skew-Hermitian operators.

Let **A** be a linear operator on a unitary space V. Then **A** is said to be

(i) a *normal operator*, if $\mathbf{A}\mathbf{A}^\circ=\mathbf{A}^\circ\mathbf{A}$:

(ii) a *Hermitian* or *self-adjoint operator*, if $\mathbf{A}^\circ=\mathbf{A}$;

(iii) a *skew-Hermitian operator*, if $\mathbf{A}^\circ=-\mathbf{A}$.

A Hermitian or a Skew-Hermitian operator is evidently a normal operator.

Referred to an orthonormal basis of V, **A** and $\mathbf{A}^\circ$ are represented by matrices A and A^{*t} (by X.4.2). Also, a matrix A is said to be normal, if $AA^{*t}=A^{*t}A$, a matrix A is said to be Hermitian, if $A^{*t}=A$, and a matrix A is said to be skew-Hermitian, if $A^{*t}=-A$. Hence it follows that :

X.5.1. *If a linear operator* **A**, *on a unitary space V, be represented by the matrix A, when referred to an orthonormal basis of V, then* **A** *is*

(i) *a normal operator if, and only if, A is a normal matrix ;*

(ii) *a Hermitian operator if, and only if, A is a Hermitian matrix ;*

(iii) *a skew-Hermitian operator if, and only if, A is a skew-Hermitian matrix.*

By virtue of the above proposition, properties of normal, Hermitian and skew-Hermitan operators follow from the properties of normal, Hermitian and skew-Hermitian matrices respectively. Thus, one obtains the following results from VIII.5.5 and VIII.5.7 :

X.5.2. (i) *The eigen values of a Hermitian operator are all real, and those of a skew-Heimitian operator are purely imaginary numbers or zero.*

(ii) *For a normal operator, eigen vectors corresponding to distinct eigen values are orthogonal (and the same is true, in particular, for Hermitian and skew-Hermitian oprators).*

Again, if a linear operator **A** on a unitary space *V* be represented by a matrix *A*, with respect to an orthonormal basis of *V*, then any matrix unitarily similar to *A* also represents the linear operator **A**, with respect to a suitable orthonormal basis of *V*. Hence, it follows from IX.5 5, IX 5 7 and IX.5.1 that :

X.5.3. *If* **A** *be a linear operator on a unitary space V, then there exists an orthonormal basis of V, with respect to which,* **A** *is represented by a diagonal matrix if, and only if,* **A** *is a normal operator.*

In particular, there exists an orthonormal basis of *V* with respect to which a Hermitian or a skew-Hermitian operator **A** is represented by a diagonal matrix (whose diagonal elements are the eigen values of **A**). Since the eigen values of a Hermitian operator are real (by X.5.2) it follows, particularly, that :

X.5.4. *Given a Hermitian operator* **A** *on a unitary space V, there exists an orthonormal basis of V, referred to which,* **A** *is representd by a real diagonal matrix (the diagonal elements being the eigen values of* **A** *).*

Non-negative Hermitian operators. A Hermitian operator **A** on a unitary space V of rank n is said to be *non-negative*, if

$$[\alpha \mathbf{A}, \alpha] \geqslant 0, \text{ for all } \alpha \text{ in } V. \qquad \dots (1)$$

If the equality sign in (1) holds only when α is the null vector, then **A** is called a *positive definite* operator (or simply a *positive* operator).

It follows directly from the above definitions that :

X.5.5. *Let V be a unitary space of rank n.*

(i) *The product of a linear operator on V and its adjoint is always a non-negative Hermitian operator.*

(ii) *All eigen values of a non-negative Hermitian operator on V are non-negative. And, conversely, if all the eigen values of a Hermitian operator on V be (real) non-negative numbers, then it is a non negative operator.*

(iii) *A non-negative Hermitian operator on V is positive definite if, and only if, it is non-singular.*

Proof : (i) Let **A** be a linear operator on V, Then,

$$(\mathbf{A}\mathbf{A}^{\circ})^{\circ} = (\mathbf{A}^{\circ})^{\circ}\mathbf{A}^{\circ} = \mathbf{A}^{\circ}\mathbf{A} \text{ and } (\mathbf{A}^{\circ}\mathbf{A})^{\circ} = \mathbf{A}^{\circ}\mathbf{A}\,;$$

also for any $\alpha \varepsilon V$,

$$[\alpha \mathbf{A}\mathbf{A}^{\circ}, \alpha] = [\alpha \mathbf{A}, \alpha \mathbf{A}] \geqslant 0, \quad [\alpha \mathbf{A}^{\circ}\mathbf{A}, \alpha] = [\alpha \mathbf{A}^{\circ}, \alpha \mathbf{A}^{\circ}] \geqslant 0.$$

Hence, both $\mathbf{A}\mathbf{A}^{\circ}$ and $\mathbf{A}^{\circ}\mathbf{A}$ are non-negative Hermitian operators.

(ii) Let **A** be a non-negative Hermitian operator on V, and let λ be an eigen value of **A**. Also, let β be an eigen vector of **A**, corrsponding to λ. Then

$$[\beta \mathbf{A}, \beta] = [\lambda\beta, \beta] = \lambda[\beta, \beta] \geqslant 0\,;$$

hence $\lambda \geqslant 0$. Thus every eigen value of **A** is non-negative.

Conversely, let **A** be a Hermitian operator on V, all whose eigen values are (real) non-negative numbers. Let $\alpha_1, \dots, \alpha_n$ form an orthonormal basis of V, relative to which **A** is represented by a real diagonal matrix (the existence of such a basis is assured by X.5.4). Then $\alpha_1,\ldots,\alpha_n$ are eigen vectors of **A** (vide IX.2.3). Let $\lambda_1,\ldots,\lambda_n$ be the eigen values of **A**, corresponding to eigen vectors $\alpha_1,\ldots,\alpha_n$ respectively.

Let $\xi = x_1\alpha_1+\ldots+x_n\alpha_n$ be an arbitrary vector in V. Then

$$\xi\mathbf{A} = (x_1\alpha_1+\ldots+x_n\alpha_n)\mathbf{A} = (x_1\lambda_1)\alpha_1+\ldots+(x_n\lambda_n)\alpha_n\ ;$$

consequently, $[\xi\mathbf{A}, \xi] = \lambda_1 x_1 x_1^* + \ldots + \lambda_n x_n x_n^* \geqslant 0,$ ... (2)

since $\lambda_i \geqslant 0$, and $x_i x_i^* = |x_i|^2 \geqslant 0$, for $i=1,\ldots,n$

The Hermitian operator **A** is, therefore, non-negative.

(iii) Let **A** be a non-negative Hermitian operator on V, and let $\lambda_1,\ldots,\lambda_n$ be the eigen values of **A**. Then, by (ii), each $\lambda_i \geqslant 0$. Also, relative to a suitable orthonormal basis of V (formed by eigen vectors of **A**), **A** is represented by the diagonal matrix $A = \text{diag}(\lambda_1,\ldots,\lambda_n)$. Thus $\det(A) = \lambda_1\ldots\lambda_n$.

If **A** be non-singular, then $\det(A) \neq 0$, so that each $\lambda_i \neq 0$; and therefore, each $\lambda_i > 0$ It then follows from (2) that $[\xi\mathbf{A}, \xi] = 0$ if, and only if, $\xi = \theta$, the null vector. Hence **A** is positive definite.

On the other hand, if $\det(A) = 0$, then one at least of the eigen values, say λ_1, must be 0. And for the corresponding eigen vector α_1, one obtains $[\alpha_1\mathbf{A}, \alpha_1] = [0.\alpha_1, \alpha_1] = 0[\alpha_1, \alpha_1] = 0$, so that **A** cannot be positive definite.

Hence **A** is positive definite if, and only if, **A** is non-singular.

It follows immediately from the above proposition that :

Corollary (i) *The square of any Hermitian operator is a non-negative Hermitian operator.*

(ii) *A non-negative Hermitian operator is non-singular if, and only if, all its eigen values are positive (real) numbers.*

X.5.6. *For each non-negative Hermitian operator* **A** *of a unitary space* V, *there exists a non-negative Hermitian operator* **B**, *such that* $\mathbf{B}^2 = \mathbf{A}$. *And, if* **A** *is non-singular, then* **B** *is also so.*

(*The relation* $\mathbf{B}^2 = \mathbf{A}$ *is also denoted by* $\mathbf{B} = \sqrt{\mathbf{A}} = \mathbf{A}^{\frac{1}{2}}$).

Proof: There exists a suitable orthonormal basis of V, with respect to which **A** is represented by the diagonal matrix,

$$A = \text{diag}\,(\lambda_1, \ldots, \lambda_n),$$

where $\lambda_1, \ldots, \lambda_n$ are the eigen values of **A**. By X.5.5(ii), all $\lambda_i \geqslant 0$; then $\sqrt{\lambda_1}, \ldots \sqrt{\lambda_n}$ are also negative real numbers.

Let $B = \text{diag}\,(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$. Then the Hermitian operator **B**, represented by the matrix B, is non-negative (by X.5.5 (ii)). Also, from $B^2 = A$, it follows that $\mathbf{B}^2 = \mathbf{A}$.

If, moreover, **A** is non-singular, then each $\lambda_i > 0$ (by Corollary (ii) of X.5.5). Then each $\sqrt{\lambda_i} > 0$, and **B** is necessarily non-singular.

Euclidean linear space. For a Euclidean linear space, the ground field is the field of real numbers **E**. Since every real number is self-conjugate, it follows from X.4.2 that :

X.5.7. *If* **A** *be a linear operator on a Euclidean linear space* V *of rank* n, *then a linear operator* **B** *on* V *is the adjoint of* **A** *if, and only if, the matrix representing* **B** *is the transpose of the matrix representing* **A**, *when referred to an orthonormal basis of* V.

A Hermitian (or self-adjoint) operator on a Euclidean linear space is called a *symmetric operator* ; and a skew-Hermitian operator on a Euclidean linear space is called a *skew-symmetric operator*.

Again, a real matrix A is called symmetric, if $A^t = A$, and it is called skew-symmetric, if $A^t = -A$. Hence, it follows from X.4.3 that :

X.5.8. *If a linear operator* **A** *on a Euclidean linear space* V *be represented by the matrix* A, *when referred to an orthonormal of* V, *then* **A** *is*

(i) *a symmetric operator if, and only if,* A *is a symmetric matrix* ;

(ii) *a skew-symmetric operator if, and only if,* A *is a skew-symmetric matrix*.

By virtue of the above proposition, properties of symmetric and skew-symmetric operators follow from the properties of symmetric and skew-symmetric matrices respectively. One obtains, in particular :

X.5.9. *For a symmetric operator* **A** *on a Euclidean linear space* V, *of rank* n,

(i) *the eigen values are all real ;*

(ii) *there exists an orthonormal basis of* V, *with respect to which,* **A** *is represented by a real diagonal matrix (the diagonal elements being the eigen values of* **A**).

The terms non-negative and positive definite are also applicable to symmetric operators on a Euclidean linear space in the same sense as in the case of Hermitian operators on a unitary space. And, the propositions X.5.5 and X.5.6 hold equally well for symmetric operators on a Euclidean linear space.

EXAMPLE 1 : If **A** be a Hermitian operator on a unitary space V, and S be the orthogonal complement of the linear subspace, generated by an eigen vector β of **A**, in V, then S is an invariant subspace of **A**.

Soln : Let λ be the eigen value of **A**, corresponding to the eigen vector β ; then $\beta A=\lambda\beta$. Let $\alpha \in S$, then $[\alpha,\beta]=0$. Now,

$$[\alpha A,\beta]=[\alpha,\beta A^{\circ}]=[\alpha,\beta A]=[\alpha,\lambda\beta]=\lambda^{*}[\alpha,\beta]=0.$$

Hence $\alpha A \in S$; and S is therefore an invariant subspace of **A**.

EXAMPLE 2 : *If* A *and* B *be two Hermitian operators on a unitary space* V, *then there exists an orthogonal basis of* V, *with respect to which* A *and* B *are represented simultaneously by two diagonal matrices if, and only if,* AB=BA.

Soln : Let AB=BA. Then, by example 8(ii), page 247, there exists a common eigen vector β_1 of A and B ; i.e.,

$$\beta_1 A=\lambda_1\beta_1, \quad \text{and} \quad \beta_1 B=\mu_1\beta_1,$$

where λ_1 and μ_1 are the eigen values of **A** and **B**, corresponding to the common eigen vector β_1.

Let S_1 be the orthogonal complement of the linear subspace, generated by β_1, in V, then S_1 is an invariant subspace of both A and B (by example 1).

Considering the linear operators A,B on S_1 (i.e., considering their induced operators A_1,B_1 on S_1), it follows as above, that there exists a vector β_2 in S_1, which is a common eigen vector of both **A** and **B**, then β_2 is orthogonal to β_1, and

$$\beta_2 A=\lambda_2\beta_2, \quad \text{and} \quad \beta_2 B=\mu_2\beta_2, \quad \text{where } \lambda_2, \mu_2 \in C.$$

Let S_2 be the orthogonal complement of the linear subspace, generated by β_2, in S_1, then S_2 is an invariant subspace of both A and B, and so on. Proceeding in this way, one obtains an orthogonal system of common eigen vectors, $\beta_1,\ldots,\beta_n$, of A and B, such that

$$\beta_i \mathrm{A}=\lambda_i\beta_i, \text{ and } \beta_i \mathrm{B}=\mu_i\beta_i, \text{ where } \lambda_i, \mu_i \varepsilon C,$$

for $i=1,\ldots,n$; then $\beta_1,\ldots,\beta_n$ form an orthogonal basis of V.

The linear operators A and B are represented by the two diagonal matrices,

$$\operatorname{diag}(\lambda_1,\ldots,\lambda_n) \text{ and } \operatorname{diag}(\mu_1,\ldots,\mu_n). \qquad \ldots (3)$$

respectively, with respect to the basis $\beta_1,\ldots,\beta_n$.

Conversely, let A and B be represented simultaneously by the two diagonal matrices (3), relative to some orthogonal basis of V. As the two diagonal matrices are commutative, the linear operators are also so, i.e., AB=BA holds.

Exercises

1. Let three linear operators A, B, C, on a Euclidean linear space V of rank 3, be represented by the following matrices A, B, C, respectively, with respect to an orthonormal basis of V,

$$A=\begin{pmatrix} 3 & -4 & 2 \\ -4 & 6 & -2 \\ 2 & -2 & 5 \end{pmatrix}, \quad B=\begin{pmatrix} 1 & -1 & 2 \\ -1 & 3 & -2 \\ 2 & -2 & 4 \end{pmatrix},$$

$$C=\begin{pmatrix} 1 & 3 & 1 \\ -1 & 2 & 2 \\ 1 & 0 & 2 \end{pmatrix}.$$

Show that A is a positive definite symmetric operator, B is a non-negative symmetric operator, and that although C is not symmetric $[\alpha\mathrm{C}, \alpha]\geqslant 0$ holds for all $\alpha \varepsilon V$.

2. Show that a linear combination of non-negative Hermitian operators (on a unitary space) with non-negative (real) cofficients is also a non-negative Hermitian operator.

3. The product of two permutable, non-negative, Hermitian operators is also a non-negative Hermitian operator.

4. Let A be a linear operator on a finite dimensional unitary space V.

(i) If A is Hermitian, and $\xi^2\mathrm{A}=\theta$, then $\xi\mathrm{A}=\theta$.

(ii) A is Hermitian if, and only if, $[\xi A, \xi]$ is real for all $\xi \varepsilon V$.

5. If A and B be two Hermitian operators, of which A is positive definite, then the eigen values of the linear operator AB are all real

If, moreover, B is non-negative, then eigen values of AB are non-negative.

6. Unitary and orthogonal operators.

Unitary oparators. A linear operator A on a unitary space V is said to be *unitary*, if it preserves the value of the inner product of every pair of vectors in V, that is,

$$[\alpha A, \beta A]=[\alpha, \beta] \text{ for every pair of vectors } \alpha, \beta \varepsilon V. \quad \ldots (1)$$

X.6.1. *Let V be a finite dimensional unitary space.*

(i) *The norm of any vector in V remains unaltered by a unitary operator on V.*

(ii) *Conversely, a linear operator on V, by which the norm of every vector in V remains unaltered, is unitary.*

Proof: (i) Let A be a unitary operator on V, and let $\alpha \varepsilon V$. Then $[\alpha A, \alpha A]=[\alpha, \alpha]$, by (1); hence $\|\alpha\| A=\|\alpha\|$.

(ii) Let A be a linear operator on V, by which norm of every vector in V remains unaltered. Thus

$$[\alpha A, \alpha A]=[\alpha, \alpha] \text{ holds, for every } \alpha \varepsilon V. \quad \ldots (2)$$

Let α, β be two arbitrary vectors in V, and let αA and βA be denoted by α' and β' respectively.

As A is linear,

$$(\alpha+b\beta)A=\alpha A+b.\beta A=\alpha'+b\beta',$$

where b is an arbitrary scalar. It now follows from (2) that

$$[\alpha+b\beta, \alpha+b\beta]=[\alpha'+b\beta', \alpha'+b\beta'],$$

i.e., $$[\alpha, \alpha]+b[\beta, \alpha]+b^*[\alpha, \beta]+bb^*[\beta, \beta]$$
$$=[\alpha', \alpha']+b[\beta', \alpha']+b^*[\alpha', \beta']+bb^*[\beta', \beta']$$

i.e., $$b[\beta, \alpha]+b^*[\alpha, \beta]=b[\beta', \alpha']+b^*[\alpha', \beta']. \quad \ldots (3)$$

Putting $b=1$ in (3), one obtains

$$[\beta, \alpha]+[\alpha, \beta]=[\beta', \alpha']+[\alpha', \beta'] \qquad \ldots (4)$$

Next, putting $b=i$ in (3), and cancelling i, one gets

$$[\beta, \alpha]-[\alpha, \beta]=[\beta', \alpha]-[\alpha', \beta']. \qquad \ldots (5)$$

From (4) and (5), it follows that

$$[\alpha, \beta]=[\alpha', \beta']=[\alpha\, \mathbf{A}, \beta \mathbf{A}]\, ; \qquad \ldots (6)$$

consequently, **A** is a unitary operator.

X.6.2. *A linear operator* **A** *on a unitary space V is unitary if, and only if,* **A** *transforms an orthonormal basis of V into an orthonormal basis.*

Proof: Let **A** be a unitary operator on V, and let $\alpha_1, \ldots, \alpha_n$ be an orthonormal basis of V. Since a unitary operator preserves the value of inner product, an orthogonal set of vectors is transformed to an orthogonal set of vectors. Also, the norm of a vector is preserved by a unitary operator ; hence unit vectors are transformed to unit vectors. Consequently, the vectors $\alpha_1\mathbf{A},\ldots,\alpha_n\mathbf{A}$ form an orthonormal system in V; and since their number is equal to the rank of V, it follows that the vectors $\alpha_1\mathbf{A},\ldots, \alpha_n\mathbf{A}$ form an orthonormal basis of V.

Conversely, let **A** be a linear operator on V, such that it transforms an orthonormal basis $\alpha_1, \ldots, \alpha_n$ of V into an orthonormal basis $\alpha_1\mathbf{A},\ldots,\alpha_n\mathbf{A}$. Let ξ and η be two arbitary vectors in V, and let $\xi=x_1\alpha_1+\ldots+x_n\alpha_n$, and $\eta=y_1\alpha_1+\ldots+y_n\alpha_n$. Then

$$[\xi, \eta]=[x_1\alpha_1+\ldots+x_n\alpha_n, y_1\alpha_1+\ldots+y_n\alpha_n]$$
$$=x_1y_1^*+\ldots+x_ny_n^*,$$
$$[\xi\mathbf{A},\eta\mathbf{A}]=[(x_1\alpha_1+\ldots+x_n\alpha_n)\mathbf{A}, (y_1\alpha_1+\ldots+y_n\alpha_n)\mathbf{A}]$$
$$=[x_1.\alpha_1\mathbf{A}+\ldots+x_n.\alpha_n\mathbf{A}, y_1.\alpha_1\mathbf{A}+\ldots+y_n.\alpha_n\mathbf{A}]$$
$$=x_1y_1^*+\ldots+x_ny_n^*,$$

so that $[\xi\mathbf{A},\eta\mathbf{A}]=[\xi,\eta]$. Hence **A** is a unitary operator on V.

Let **A** be a unitary operator on a unitary space V. Then for every pair of vectors $\alpha, \beta \in V$, $[\alpha, \beta]=[\alpha\mathbf{A},\beta\mathbf{A}]$. But

$$[\alpha\mathbf{A},\beta\mathbf{A}]=[\alpha,\beta\mathbf{A}\mathbf{A}^\circ], \text{ so that } [\alpha,\beta]=[\alpha,\beta\mathbf{A}\mathbf{A}^\circ].$$

Hence $$\mathbf{A}\mathbf{A}^\circ=\mathbf{I}, \text{ i.e.. } \mathbf{A}^\circ=\mathbf{A}^{-1}. \qquad \ldots (7)$$

Conversely, if $\mathbf{A}^{\circ}=\mathbf{A}^{-1}$, then it follows, by reversing the above arguments, that **A** must be a unitary operator. Thus one obtains :

X.6.3. *A linear operator* **A** *on a unitary space* V *is unitary if, and only if, its adjoint* $\mathbf{A}^{\circ}$ *is the inverse of* **A**.

Let **A** be a linear operator on a unitary space V. Let **A** be represented by the matrix A, with respect to an orthonormal basis of V. Then $\mathbf{A}^{\circ}$ is represented by the matrix A^{*t}. Hence if **A** be a unitary operator, it follows from (7) that $AA^{*t}=I$, so that A is a unitary matrix.

Conversely, if A be a unitary matrix, then it follows by reversing the above arguments that **A** must be a unitary operator. Thus one obtains the following :

X.6.4. *A linear operator on a unitary space* V *is unitary if, and only if, it is represented by a unitary matrix, when referred to an orthonormal basis of* V.

The above proposition enables one to derive properties of unitary operators from the properties of unitary matrices.

Of particular interest is the following proposition, which gives a *polar decomposition* of a linear operator :

X.6.5. *Every non-singular linear operator* **A** *on a unitary space* V *admits a polar decomposition,*

$$\mathbf{A}=\mathbf{H}\mathbf{U},$$

where **H** *is a non-singular, non-negative Hermitian operator, and* **U** *a unitary operator.*

Proof: By X.5.5.(i), $\mathbf{A}\mathbf{A}^{\circ}$ is a non-negative Hermitian operator. Hence, by X.5.6, there exists a non-singular, non-negative Hermitian operator **H**, satisfying $\mathbf{H}^{2}=\mathbf{A}\mathbf{A}^{\circ}$.

Let $\mathbf{U}=\mathbf{H}^{-1}\mathbf{A}$, then **U** is unitary. In fact,

$$\mathbf{U}\mathbf{U}^{\circ}=(\mathbf{H}^{-1}\mathbf{A})(\mathbf{H}^{-1}\mathbf{A})^{\circ}=\mathbf{H}^{-1}(\mathbf{A}\mathbf{A}^{\circ})\mathbf{H}^{-1}=\mathbf{H}^{-1}\mathbf{H}^{2}\mathbf{H}^{-1}=\mathbf{I},$$

since $(\mathbf{H}^{-1})^{\circ}=(\mathbf{H}^{\circ})^{-1}=\mathbf{H}^{-1}$ (as **H** is Hermitian).

Thus $\mathbf{A}=\mathbf{H}\mathbf{U}$, where **H**, **U** satisfy the requisite conditions.

Unitary operators are very closely connected with the Hermitian and skew Hermitian operators. This can be clearly seen from the following two propositions known as *Cayley's formulae* :

X.6.6. *If* A *be a Hermitian operator on a unitary space, then*

(*i*) *the operators* $A \pm iI$ *are invertible* :

(*ii*) *the operator* U, *given by*

$$U = (A - iI)(A + iI)^{-1}, \qquad \ldots (8)$$

is unitary, and does not have 1 *as an eigen value* ;

(*iii*) A *is expressed in terms of* U *by*

$$A = -i(U + I)(U - I)^{-1}. \qquad \ldots (9)$$

Conversely, if U *be a unitary operator, which does not have* 1 *as an eigen value, then*

(*iv*) *the operator* $U - I$ *is invertible* ;

(*v*) *the operator* A, *given by* (9), *is Hermitian* ;

(*vi*) U *is expressed in terms of* A *by* (8).

Proof : Let A be a Hermitian operator on a unitary space V.

(i) Since the eigen values of A are real, $-i$ or i cannot be an eigen value of A. Hence the operators $A \pm iI$ are non-singular.

(ii) $(A + iI)(A - iI) = A^2 + iA - iA + I = (A - iI)(A + iI)$.

Thus $A + iI$ is permutable with $A - iI$; hence $A + iI$ is also permutable with $(A - iI)^{-1}$. Now, by (8),

$$\begin{aligned} UU^{\circ} &= [(A - iI)(A + iI)^{-1}][(A - iI)(A + iI)^{-1}]^{\circ} \\ &= (A - iI)(A + iI)^{-1}.(A - iI)^{-1}(A + iI) \\ &= (A - iI)(A + iI)^{-1}.(A + iI)(A - iI)^{-1} = I, \end{aligned}$$

since $[(A - iI)(A + iI)^{-1}]^{\circ} = [(A + iI)^{\circ}]^{-1}.(A - iI)^{\circ}$

$$= (A - iI)^{-1}(A + iI).$$

Thus U is a unitary operator.

Again, from (8), one has

$$\begin{aligned} (U - I)(A + iI) &= [(A - iI)(A + iI)^{-1} - I](A + iI) \\ &= (A - iI) - (A + iI) = -2iI, \qquad \ldots (10) \end{aligned}$$

so that $U-I$ is invertible. Hence 1 cannot be an eigen value of U.

(iii) It follows from (10) that

$$(U-I)A=-2iI-i(U-I)=-i(U+I),$$

so that $$A=-i(U-I)^{-1}(U+I)=-i(U+I)(U-I)^{-1}.$$

Conversly, let U be a unitary operator, which does not have 1 as an eigen value.

(iv) Since 1 is not an eigen value of U, it follows that $U-I$ is invertible.

(v) Let A be an operator given by (9), then

$$\begin{aligned}A^{\circ}&=[-i(U+I)(U-I)^{-1}]^{c}=i\,[(U-I)^{\circ}]^{-1}(U+I)^{\circ}\\&=i\,(U^{\circ}-I)^{-1}(U^{\circ}+I)=i(U^{-1}-I)^{-1}(U^{-1}+I)\\&=i(I-U)^{-1}U.U^{-1}(I+U)=-i(U+I)(U-I)^{-1}=A.\end{aligned}$$

Hence A is a Hermitian operator.

(vi) The relation (8) can be deduced from the relation (9), by reversing the arguments in the proofs of (iii) and (ii) above.

By formulae (8) and (9) a one-to-one correspondence is established between the Hermitian operators and those unitary operators, which do not have 1 as an eigen value. Similarly, a one-to-one correspondence between the Hermitian operators and those unitary operators, which do not have -1 as an eigen value, can be established by the formulae :

$$U=(iI+A)(iI-A)^{-1}, \qquad \ldots (11)$$

and $$A=i(U-I)(U+I)^{-1}. \qquad \ldots (12)$$

X.6.7. *If A be a skew-Hermitian operator on a unitary space, then*

(i) *the operators* $A\pm I$ *are invertible ;*

(ii) *the operator U, given by*

$$U=(A-I)(A+I)^{-1}, \qquad \ldots (13)$$

is unitary, and does not have 1 as an eigen value ;

(iii) *A is expressed in terms of U by*

$$A=-(U+I)(U-I)^{-1}. \qquad \ldots (14)$$

Conversely, if U *be a unitary operator, which does not have* 1 *as an eigen value, then*

(iv) *the operator* $U-I$ *is invertible ;*

(v) *the operator* A, *given by* (14), *is skew-Hermitian ;*

(vi) U *is expressed in terms of* A *by* (13).

Proof : Let A be a skew-Hermitian operator on a unitary space *V*.

(i) Since the eigen values of A are purely imaginary numbers or zero, 1 or -1 cannot be an eigen value of A. Hence the operators $A \pm I$ are invertible.

(ii) $A+I$ is permutable with $A-I$, and hence also with $(A-I)^{-1}$.

Now,
$$UU^{\circ}=[(A-I)(A+I)^{-1}][(A-I)(A+I)^{-1}]^{\circ}$$
$$=(A-I)(A+I)^{-1}.(A^{\circ}+I)^{-1}(A^{\circ}-I)$$
$$=(A-I)(A+I)^{-1}(-A+I)^{-1}(-A-I)$$
$$=(A-I)(A+I)^{-1}(A-I)^{-1}(A+I)=I.$$

Thus U is a unitary operator.

Again, from (13), it follows that
$$(U-I)(A+I)=[(A-I)(A+I)^{-1}-I](A+I)$$
$$=(A-I)-(A+I)=-2I, \qquad \ldots (15)$$

so that $U-I$ is invertible. Hence 1 cannot be an eigen value of U.

(iii) It follows from (15) that
$$(U-I)A=-2I-(U-I)=-(U+I),$$
so that
$$A=-(U+I)(U-I)^{-1}.$$

Conversely, let U be a unitary operator, which does not have 1 as an eigen value.

(iv) Since 1 is not an eigen value of U, it follows that $U-I$ is invertible.

(v) Let A be an operator given by (14), then
$$A^{\circ}=[(U+I)(U-I)^{-1}]^{\circ}=(U^{\circ}-I)^{-1}(U^{\circ}+I)$$
$$=(U^{-1}-I)^{-1}(U^{-1}+I)=(I-U)^{-1}(I+U)$$
$$=-(U+I)(U-I)^{-1}=-A.$$

Hence A is a skew-Hermitian operator.

(vi) The relaltion (13) can be deduced from the relation (14), by reversing the arguments in the proofs of (iii) and (ii) above.

By the formulae (13) and (14), a one-to-one correspondence has been established between the skew-Hermitian operators and those unitary operators, which do not have 1 as an eigen value. Similarly, a one-to-one correspondence between the skew-Hermitian operators and those unitary operators, which do not have -1 as an eigen value, can be established by the formulae :

$$\mathbf{U}=(\mathbf{I}+\mathbf{A})(\mathbf{I}-\mathbf{A})^{-1}, \qquad \ldots(16)$$

and

$$\mathbf{A}=-(\mathbf{I}-\mathbf{U})(\mathbf{I}+\mathbf{U})^{-1}. \qquad \ldots(17)$$

Orthogonal operators. A unitary operator on a Euclidean linear space is called an *orthogonal operator.* Thus a linear operator $\mathbf{A}$ on a Euclidean linear space V is said to be orthogonal, if it preserves the value of the inner product of every pair of vectors in V, i,e.,

$$[\alpha\mathbf{A}, \beta\mathbf{A}]=[\alpha, \beta], \text{ for every pair of vectors } \alpha, \beta \in V.$$

The properties, proved in propositions X.6.1 to X.6.5 and X.6.7, hold also for orthogonal operators on a Euclidean linear space with necessary modifications. The analogues of the above propositions are :

X.6.8. *Let V be a finite dimensional Euclidean linear space.*

(i) *The norm of any vector in V remains unaltered by an orthogonal operator on V.*

(ii) *Conversely, a linear operator on V, by which the norm of every vector in V remains unaltered, is an orthogonal operator.*

Proof: The proof of (i) is the same as in X.6.1. The proof of (ii) is the same as in X.6.1, with the modification that (6) follows directly from (4), since in the present case $[\alpha,\beta]=[\beta,\alpha]$ and $[\alpha',\beta']=[\beta',\alpha']$. (Thus the scalar i need not be used).

X.6.9. *A linear operator $\mathbf{A}$ on a Euclidean linear space V is orthogonal if, and only if, $\mathbf{A}$ transforms an orthonormal basis of V into an orthonormal basis.*

Proof: The proof is the same as in X.6.?.

X.6.10. *A linear operator $\mathbf{A}$ on a Euclidean linear space V is orthogonal if, and only if, its adjoint $\mathbf{A}^{\circ}$ is the inverse of $\mathbf{A}$.*

Proof : The same as for X.6.3.

X.6.11. *A linear operator on a Euclidean linear space V is orthogonal if, and only if, if it is represented by an orthogonal matrix, when referred to an orthonormal basis of V.*

Proof : The proof is similar to that for X.6.4.

X.6.12. *Every non-singular linear operator* $\mathbf{A}$ *on a Euclidean linear space V admits a polar decomposition.*

$$\mathbf{A}=\mathbf{S}\mathbf{G},$$

where $\mathbf{S}$ *is a non-singular, non negative, symmetric operator, and* $\mathbf{G}$ *is an orthogonal operator.*

Proof : Similar to that for X.6.5.

X.6.13. *If* $\mathbf{A}$ *be a skew-symmetric operator on a Euclidean linear space, then*

(i) *the operators* $\mathbf{A}\pm\mathbf{I}$ *are invertible ;*

(ii) *the operator* $\mathbf{G}$*, given by*

$$\mathbf{G}=(\mathbf{A}-\mathbf{I})(\mathbf{A}+\mathbf{I})^{-1} \qquad \dots (18)$$

is orthogonal, and does not have 1 *as an eigen value ;*

(iii) $\mathbf{A}$ *is expressed in terms of* $\mathbf{G}$ *by*

$$\mathbf{A}=-(\mathbf{G}+\mathbf{I})(\mathbf{G}+\mathbf{I})^{-1}. \qquad \dots (19)$$

Conversely, if $\mathbf{G}$ *be an orthogonal operator, which does not have* 1 *as an eigen value, then*

(iv) *the operator* $\mathbf{G}-\mathbf{I}$ *is invertible,*

(v) *the operator* $\mathbf{A}$ *given by* (19), *is skew-Hermitian,*

(vi) $\mathbf{G}$ *is expressed in terms of* $\mathbf{A}$ *by* (18).

Proof : The proof is similar to that for X 6.7.

Similar relations between skew-symmetric operators and those orthogonal operators, which do not have -1 as an eigen value, are given by the formulae :

$$\mathbf{G}=(\mathbf{I}-\mathbf{A})(\mathbf{I}+\mathbf{A})^{-1}, \qquad \dots (20)$$

and

$$\mathbf{A}=(\mathbf{I}-\mathbf{G})(\mathbf{I}+\mathbf{G})^{-1}, \qquad \dots (21)$$

where $\mathbf{A}$ is skew-symmetric and $\mathbf{G}$ is orthogonal.

Exercises

1. If the linear operators A, B on a unitary space V change the norm of vectors in the same way, that is, if

$$[\alpha A, \alpha A]=[\alpha B, \alpha B], \quad \text{for every vector } \alpha \in V,$$

then there exists a unitary operator U on V, such that $B=AU$.

2. Apply exercise 1 to generalise the proposition X.6.5 to cover the case of linear operators A, which may not be non-singular; and show that H is unique, and that U is also unique provided that A is non-singular.

3. If λ be an eigen value of A, then $(1-i\lambda)(1+i\lambda)^{-1}$ is an an eigen value of U, where A and U are operators defined in X.6.6(ii).

4. Establish the formulae (11) and (12), (16) and (17), and (20) and (21) under the conditions stated for them.

5. If A and B be Hermitian operators on a unitary space V, then A is permutable with B if, and only if, A and B are polynomials, with real coefficients, in a common Hermitian operator C on V.

7. Spectral decomposition.

Projection operators. Let V be a unitary space of rank n. Let S be a linear subspace of V, and T be the orthogonal complement of S in V. Then V is the direct sum of S and T, and every vector $\alpha \varepsilon V$ is expressible as a sum

$$\alpha=\sigma+\tau, \qquad \sigma \varepsilon S, \ \tau \varepsilon T, \qquad \dots (1)$$

where σ and τ are uniquely determined by the relation (1). [Then σ is called the orthogonal projection of α on S, and τ is called the component of α orthogonal to S (vide page 50)].

Let P_S be an operator on V, defined by

$\alpha P_S=\sigma$, where σ is the orthogonal projection of α on S; then P_S is called a *projection operator* on V, mapping V onto S. For a fixed linear subspace S, the projection operator P_S may, for the sake of simplicity, be denoted by P (only).

Some properties of this operator are given in the following proposition :

X.7.1. *If* **P** *be the projection operator on the unitary space* V, *mapping* V *onto a linear subspace* S, *then*

(i) *the operator* **P** *is linear* ;

(ii) *the operator* **P** *is non-negative symmetric* ;

(iii) *the operator* **P** *is idempotent.*

Proof : Let α, β be two arbitrary vectors in V, having the unique decompositions,

$$\alpha=\sigma+\tau,\quad \beta=\sigma'+\tau',\ \text{where } \sigma, \sigma' \varepsilon S \text{ and } \tau, \tau' \varepsilon T,$$

so that $\alpha P=\sigma$ and $\beta P=\sigma'$.

(i) Let b be an arbitrary scalar. Then one obtains the unique decompositions,

$$\alpha+\alpha'=(\sigma+\sigma')+(\tau+\tau'),\ \text{and}\ b\alpha=(b\sigma)+(b\tau),$$

where $\sigma+\sigma'$, $b\sigma \varepsilon S$ and $\tau+\tau'$, $b\tau \varepsilon T$. Consequently,

$$(\alpha+\beta)P=\sigma+\sigma'=\alpha P+\beta P,\ \text{and}\ (b\alpha)P=b\sigma=b(\alpha P)\ ;$$

hence the operator **P** is linear.

(ii) For a linear operator **P**, $[\alpha P, \beta]=[\alpha, \beta P^{\circ}]$. Also,

$$[\alpha P, \beta]=[\sigma, \sigma'+\tau']=[\sigma, \sigma']=[\sigma+\tau, \sigma']=[\alpha, \beta P].$$

Thus $P=P^{\circ}$; so the linear operator **P** is symmetric. Also,

$$[\alpha P, \alpha]=[\sigma, \sigma+\tau]=[\sigma, \sigma]\geqslant 0\ ;$$

hence the symmetric operator **P** is non-negative.

(iii) For every vector $\sigma \varepsilon S \subset V$, $\sigma P=\sigma$. Hence,

$$\alpha P^2=(\alpha P)P=\sigma P=\sigma=\alpha P\ ;$$

and since the above holds for every $\alpha \varepsilon V$, it follows that $P^2=P$. Hence the linear operator **P** is idempotent.

Conversely,

X.7.2. *Every symmetric idempotent linear operator* **P** *on a unitary space* V *is a projection operator, mapping* V *onto a suitable linear subspace of* V.

Proof : Let the range of **P** be denoted by S, i.e., $VP=S$; then S is a linear subspace of V. For any vector $\alpha \varepsilon V$,

$$\alpha=\alpha P+\alpha(I-P), \qquad \ldots (2)$$

where **I** is the identity operator.

Now, $\alpha P \in VP = S$. So it will be sufficient to show that $\alpha(I-P) \in T$, where T is the orthogonal complement of S in V. Let σ' be an arbitrary vector in S, then $\sigma' = \beta P$, for some β in V (since $S = VP$). Now,

$$[\sigma', \alpha(I-P)] = [\beta P, \alpha(I-P)] = [\beta, \alpha(I-P)P^{\circ}] = [\beta, \alpha(I-P)P]$$
$$= [\beta, (\alpha - \alpha P)P] = [\beta, \alpha P - \alpha P] = [\beta, \theta] = 0\ ;$$

hence $\alpha(I-P) \in T$. Consequently, it follows from the decomposition (2) that P is a projection operator on V, mapping V onto S (i.e., onto the range of P).

X.7.3. *An operator* P *on a unitary space V, maping V onto a linear subspace* S, *is a projection operator if, and only if, it is represented by a daigonal matrix of the form,*

$$P = diag\ (1,\ldots,1,\ 0,\ldots,0), \qquad \ldots (3)$$

when referred to a suitable (orthonormal) basis of V.

Proof: Let P be a projection operator. Let T be the orthogonal complement of S in V. Then V is the direct sum of S and T ; and one can obtain an orthonormal basis of V, consisting of the vectors, $\sigma_1,\ldots,\sigma_r, \tau_1, \ldots, \tau_s$, where $\sigma_1, \ldots, \sigma_r$ form an orthonormal basis of S, and $\tau_1,\ldots,\tau_s$ form an orthonormal basis of T (vide II.3.19(i)). Then

$$\sigma_i P = \sigma_i, \text{ for } i = 1,\ldots,r, \text{ and } \tau_i P = \theta, \text{ for } i = 1,\ldots,s.$$

Hence, it follows that, referred to the above basis of V, P is represented by a matrix P of the form (3).

Conversely, if P be an operator on V, represented by a matrix P of the form (3), then as the matrix P is symmetric and idempotent, it follows that the same is true for the operator P. Hence P is a projection operator, by X.7.2.

Let P and Q be two projection operators on a unitary space V, mapping V onto the linear subspaces S_1 and S_2 respectively. P and Q are said to be *orthogonal* if $PQ = O$, where O is the null operator. Since P and Q are symmetric, it follows that

$$QP = Q^{\circ}P^{\circ} = (PQ)^{\circ}.$$

Hence $PQ = O$ implies $QP = O$; in other words, *if* P *and* Q *be orthogonal, then* Q *and* P *are also orthogonal.*

Let **P** and **Q** be orthogonal, and let $\sigma_1 \epsilon S_1$ and $\sigma_2 \epsilon S_2$. Then

$$[\sigma_1, \sigma_2] = [\sigma_1 \mathbf{P}, \sigma_2 \mathbf{Q}] = [\sigma_1, \sigma_2 \mathbf{Q} \mathbf{P}^{\circ}] = [\sigma_1, \sigma_2 \mathbf{Q} \mathbf{P}]$$
$$= [\sigma_1, \theta] = 0.$$

Thus the linear subspaces S_1 are S_2 are orthogonal.

Conversly, let S_1 and S_2 be orthogonal. Then, for an arbitrary vector $\alpha \epsilon V$.

$$\alpha(\mathbf{PQ}) = (\alpha \mathbf{P})\mathbf{Q} = \theta, \quad \text{since } \alpha \mathbf{P} \epsilon S_1,$$

so that $\mathbf{PQ} = \mathbf{O}$. Hence **P** and **Q** are orthogonal.

Thus one obtains the following proposition :

X.7.4. *If* **P** *and* **Q** *be two projection operators on a unitary space* V, *mapping* V *onto the linear subspaces* S_1 *and* S_2 *respectively, then* **P** *and* **Q** *are orthogonal if, and only if,* S_1 *and* S_2 *are orthogonal.*

Spectral decomposition Let **A** be a linear operator on a finite dimensional unitary space V.

A decomposition of **A** as a linear combination,

$$\mathbf{A} = \lambda_1 \mathbf{P}_1 + \ldots + \lambda_r \mathbf{P}_r, \qquad \ldots(4)$$

where $\lambda_1, \ldots, \lambda_r$ are distinct scalars, and $\mathbf{P}_1, \ldots, \mathbf{P}_r$ are pair-wise orthogonal, non-null, projection operators on V, such that $\mathbf{P}_1 + \ldots + \mathbf{P}_r = \mathbf{I}$, is called a *spectral decomposition* of **A**. The set of the coefficients, $\lambda_1, \ldots, \lambda_r$, is said to be the *spectrum* of the decomposition or of the linear operator **A**.

It follows from the definition that (4) is a spectral decomposition of the linear operator **A**, if the following conditions hold :

(a) the scalars $\lambda_1, \ldots \lambda_r$ are distinct ;
(b) $\mathbf{P}_i^{\circ} = \mathbf{P}_i \neq \mathbf{O}$, for $i = 1, \ldots, r$;
(c) $\mathbf{P}_i^2 = \mathbf{P}_i$, for $i = 1, \ldots, r$;
(d) $\mathbf{P}_i \mathbf{P}_j = \mathbf{O}$, for $i, j = 1, \ldots, r$, where $i \neq j$;
(e) $\mathbf{P}_1 + \ldots + \mathbf{P}_r = \mathbf{I}$.

X.7.5 *Let* $\mathbf{A} = \lambda_1 \mathbf{P}_1 + \ldots + \lambda_r \mathbf{P}_r$ *be a spectral decomposition of a linear operator* **A** *on a finite dimensional unitary space* V. *Then*

(i) *the spectrum* $[\lambda_1,\dots,\lambda_r]$ *is the set of all the distinct eigen values of* $\mathbf{A}$:

(ii) $f(\mathbf{A})=f(\lambda_1)\mathbf{P}_1+\dots+f(\lambda_r)\mathbf{P}_r$, *where* $f(x)$ *is any polynomial over the field of complex numbers* ;

(iii) $\mathbf{P}_i=t_i(\mathbf{A}-\lambda_1\mathbf{I})\dots(\mathbf{A}-\lambda_{i-1}\mathbf{I})(\mathbf{A}-\lambda_{i+1}\mathbf{I})\dots(\mathbf{A}-\lambda_r\mathbf{I})$,
where $t_i=[(\lambda_i-\lambda_1)\dots(\lambda_i-\lambda_{i-1})(\lambda_i-\lambda_{i+1})\dots(\lambda_i-\lambda_r)]^{-1}$,
for $i=1,\dots,r$.

Proof: (i) Since $\mathbf{P}_i$ is a non-null operator on V, there exists a vector $\alpha\,\varepsilon\,V$, such that $\alpha\mathbf{P}_i\neq\theta$. Then

$$(\alpha\mathbf{P}_i)\mathbf{A}=(\alpha\mathbf{P}_i)(\lambda_1\mathbf{P}_1+\dots+\lambda_r\mathbf{P}_r)=\lambda_i(\alpha\mathbf{P}_i),$$

since $\mathbf{P}_i\mathbf{P}_j=\mathbf{O}$, for $j\neq i$, and $\mathbf{P}_i^2=\mathbf{P}_i$. Hence $\alpha\mathbf{P}_i$ is an eigen vector of $\mathbf{A}$, having λ_i as the corresponding eigen value. Thus λ_i is an eigen value of $\mathbf{A}$. As the above is true for $i=1,\dots,r$, it follows that each of $\lambda_1,\dots,\lambda_r$ is an eigen value of A.

Next, let μ be an eigen value of $\mathbf{A}$, and let β be an eigen vector of $\mathbf{A}$, corresponding to μ. Then

$$\beta\mathbf{A}=\mu\beta=\mu\beta\mathbf{I}=\mu\beta\,(\mathbf{P}_1+\dots+\mathbf{P}_r)=\mu\beta\mathbf{P}_1+\dots+\mu\beta\mathbf{P}_r.$$

Again, $\beta\mathbf{A}=\beta(\lambda_1\mathbf{P}_1+\dots+\lambda_r\mathbf{P}_r)=\lambda_1\beta\mathbf{P}_1+\dots+\lambda_r\beta\mathbf{P}_r.$

Thus, $\lambda_1\beta\mathbf{P}_1+\dots+\lambda_r\beta\mathbf{P}_r=\mu\beta\mathbf{P}_1+\dots+\mu\beta\mathbf{P}_r.$

Now, operating on both sides by $\mathbf{P}_i$, one obtains, by (c) and (d),

$$\lambda_i\beta\mathbf{P}_i=\mu\beta\mathbf{P}_i,\ \text{ i.e., }\ (\lambda_i-\mu)\beta\mathbf{P}_i=\theta,\ \text{for } i=1,\dots,r.$$

Again, since β is a non-null vector, and $\beta=\beta(\mathbf{P}_1+\dots+\mathbf{P}_r)=\beta\mathbf{P}_1+\dots+\beta\mathbf{P}_r$, it follows that one at least of the summands, $\beta\mathbf{P}_1,\dots,\beta\mathbf{P}_r$ must be non-null, say $\beta\mathbf{P}_k\neq\theta$. Then $(\lambda_k-\mu)\beta\mathbf{P}_k=\theta$ implies that $\lambda_k-\mu=0$, i.e., $\mu=\lambda_k$. Thus any eigen value μ of A must be one among $\lambda_1,\dots,\lambda_r$. Hence the spectrum $[\lambda_1,\dots,\lambda_r]$ is the set of all the distinct eigen values of A.

(ii) Let $f(x)=a_0+a_1x+\dots+a_mx^m$. By (c) and (d), one obtains successively :

$$\begin{aligned}
\mathbf{I}&=\mathbf{P}_1+\dots+\mathbf{P}_r,\\
\mathbf{A}&=\lambda_1\mathbf{P}_1+\dots+\lambda_r\mathbf{P}_r,\\
\mathbf{A}^2&=(\lambda_1\mathbf{P}_1+\dots+\lambda_r\mathbf{P}_r)(\lambda_1\mathbf{P}_1+\dots+\lambda_r\mathbf{P}_r)\\
&=\lambda_1^2\mathbf{P}_1+\dots+\lambda_r^2\mathbf{P}_r,\\
&\dots\quad\dots\quad\dots\quad\dots\quad\dots\quad\dots\quad\dots\quad\dots
\end{aligned}$$

$$\begin{aligned}A^m &= (\lambda_1{}^{m-1}P_1+\ldots+\lambda_r{}^{m-1}P_r)(\lambda_1P_1+\ldots+\lambda_rP_r)\\ &= \lambda_1{}^mP_1+\ldots+\lambda_r{}^mP_r.\end{aligned}$$

Multiplying the above equalities successively by $a_0, a_1, \ldots, a_m$, and then adding these, one obtains

$$\begin{aligned}a_0I+a_1A+\ldots+a_mA^m = (a_0+a_1\lambda_1+\ldots+a_m\lambda_1{}^m)P_1+\ldots\\ +(a_0+a_1\lambda_r+\ldots+a_m\lambda_r{}^m)P_r,\end{aligned}$$

i.e.,
$$f(A)=f(\lambda_1)P_1+\ldots+f(\lambda_r)P_r.$$

(iii) Let $g_i(x)$ be the polynomial, with complex coefficients, given by $g_i(x)=t_i(x-\lambda_1)\ .\ (x-\lambda_{i-1})\ (x-\lambda_{i+1})\ldots(x-\lambda_r)$, for $i=1, \ldots, r$. Then, by (ii), one obtains

$$g_i(A)=g_i(\lambda_1)P_1+\ldots+g_i(\lambda_r)P_r=P_i,$$

since $g_i(\lambda_j)=0$ when $j=i$, and $g_i(\lambda_i)=1$. Thus

$$P_i=g_i(A)=t_i(A-\lambda_1I)\ldots(A-\lambda_{i-1}I)(A-\lambda_{i+1}I)\ldots(A-\lambda_rI),$$

for $i=1,\ldots r$.

It follows from X.7.5 (i), (iii) that *a spectral decomposition for a linear operator is necssarily unique.*

It has not yet been seen whether there exists any linear operator which admits a spectral decomposition. Existence of such operators will be proved in the following proposition.

X.7.6. *There exists a spectral decomposition for a linear operator* A, *on a unitary space* V, *if, and only if,* A *is a normal operator.*

Proof: Let A be a linear operator, for which there exists a spectral decomposition

$$A=\lambda_1P_1+\ldots+\lambda_rP_r.$$

Then $A^\circ=\lambda_1{}^*P_1{}^\circ+\ldots+\lambda^*{}_rP_r{}^\circ=\lambda_1{}^*P_1+\ldots+\lambda_r{}^*P_r$ (by(b)),

so that
$$\begin{aligned}AA^\circ &= (\lambda_1P_1+\ldots+\lambda_rP_r)(\lambda_1{}^*P_1+\ldots+\lambda_r{}^*P_r)\\ &= \lambda_1\lambda_1{}^*P_1+\ldots+\lambda_r\lambda_r{}^*P_r\\ &= A^\circ A, \qquad \text{(by (c) and (d))}.\end{aligned}$$

Thus A is necessarily a normal operator.

Conversely, let A be a normal operator. Let $\lambda_1,\ldots,\lambda_r$ be the distinct eigen values of A, and let $S_1,\ldots,S_r$ be the characteristic subspaces corresponding to $\lambda_1,\ldots,\lambda_r$ respectively. Then $S_1,\ldots,S_r$ are pair-wise orthogonal linear subspaces of V, and V is the direct sum of the subspaces $S_1,\ldots,S_r$.

Let P_i denote the projection operator on V, which maps V into S_i, for $i=1,\ldots,r$. Then every P_i is symmetric and idempotent (by X.7.1 (ii), (iii)); thus (b) and (c) hold. Also, since the linear subspaces $S_1,\ldots, S_r$ are pair-wise orthogonal, it follows (by X.7.4) that the operators $P_1,\ldots,P_r$ are pair-wise orthogonal; thus (d) holds. Finally,

$$\alpha=\sigma_1+\ldots+\sigma_r, \text{ where } \sigma_j \varepsilon S_j, \text{ for } i=1,\ldots,r.$$

Thus $\alpha I=\alpha P_1+\ldots+\alpha P_r=\alpha(P_1+\ldots+P_r)$ holds, for every vector $\alpha \varepsilon V$. Consequently $I=P_1+\ldots+P_r$, which proves the property (e).

It remains to show that A can be expressed as a linear combination,

$$A=\lambda_1 P_1+\ldots+\lambda_r P_r.$$

Let α be an arbitrary vector in V, then

$$\alpha=\alpha I=\alpha(P_1+\ldots+P_r)=\alpha P_1+\ldots+\alpha P_r$$

i.e., $$\alpha A=(\alpha P_1)A+\ldots+(\alpha P_r)A=\lambda_1(\alpha P_1)+\ldots+\lambda_r(\alpha P_r).$$

since αP_i is an eigen vector of A corresponding to the eigen value λ_i (as $\alpha P_i \varepsilon S_i$), for $i=1,\ldots,r$. Thus

$$\alpha A=\alpha(\lambda_1 P_1+\ldots+\lambda_r P_r)$$

holds, for every $\alpha \varepsilon V$. Consequently,

$$A=\lambda_1 P_1+\ldots+\lambda_r P_r,$$

where $\lambda, \ldots,\lambda_r$ are distinct (i.e., the property (a) holds).

Thus there exists a spectral decomposition for the normal operator A.

Proposition X.7.6 is some times stated in the following alternative form :

X.7.7. Spectral theorem. *If* A *be a linear operator on a finite dimensional unitary space* V, *then* V *has an orthonormal basis consisting of eigen vectors of* A *if, and only if,* A *is normal.*

It now follows immediately, by IX.5.5, that :

X.7.8. *A linear operator* A *on a finite dimensional unitary space* V *admits a spectral decomposition if, and only if,* A *is represented by a diagonal matrix, when referred to a suitable orthonormal basis of* V.

X.7.9. *A normal operator* A *on a unitay space space* V *is*

(*i*) *Hermitian if, and only if, its spectrum consists of real numbers ;*

(ii) *skew-Hermitian if, and only if, its spectrum consists of purely imaginary numbers or the number zero ;*

(iii) *unitary if, and only if, the spectrum consists of numbers with modulus* 1.

Proof : As A is a normal operator, it has a unique spectral decomposition,

$$A=\lambda_1 P_1+\ldots+\lambda_r P_r.$$

Then $$A^{\circ}=\lambda_1^* P_1+\ldots+\lambda_r^* P_r \qquad (\text{since } P_i^{\circ}=P_i).$$

(i) A is Hermitian if, and only if, $A^{\circ}=A$, i.e., $\lambda_i^*=\lambda_i$, i.e., λ_i is a real number, for $i=1,\ldots,r$.

(ii) A is skew-Hermitian if, and only if, $A^{\circ}=-A$, i.e., $\lambda_i^*=-\lambda_i$, i e., λ_i is a purely imaginary number or 0, for $i=1,\ldots,r$.

(iii) A is unitary if, and only if, $AA^{\circ}=I$,

i.e., $$(\lambda_1 P_1+\ldots+\lambda_r P_r)(\lambda_1^* P_1+\ldots+\lambda_r^* P_r)=I,$$

i.e., $$\lambda_1\lambda_1^* P_1+\ldots+\lambda_r\lambda_r^* P_r=I,$$

i.e., $\lambda_i\lambda_i^*=1$, i.e., modulus of $\lambda_i=1$, for $i=1,\ldots,r$.

Exercises

1. Let P and Q be two projection operators on a unitary space V, mapping V onto the linear subspaces S and T respectively. Then

(i) P+Q is a projection operator, mapping V onto some linear subspace W if, and only if, P and Q are orthogonal ; and in this case W is the linear sum of S and T ;

(ii) PQ is a projection operator, mapping V onto some linear subspace G, if, and only if, P and Q are permutable ; and in this case $G = S \cap T$,

(iii) $S \subset T$ if, and only if, PQ=P ;

(iv) if T be the orthogonal complement of S in V, then Q=I−P.

2. Let V be a finite dimensional linear space, over a field K. Then

(i) Every projection operator on V, except the identity operator, is singular.

(ii) A vector remains fixed under a projection operator P if, and only if, it belongs to the range of P.

(iii) An operator A on V is a projection operator if, and only if, the eigen values of A are *1* and *0*.

(iv) V is the direct sum of the null space and the range of any projection operator on V.

(v) If V is the linear sum of the null space and the fixed space of a linear operator A on V, then A is a projection operator. [The fixed space of A is the linear subspace formed by all those vectors in V which remain fixed under A.]

3. If A=HU be the polar decomposition (vide X 6.5) of a *normal* operator A, then

$$\mathrm{HU}=\mathrm{UH},\ \mathrm{AU}=\mathrm{UA} \text{ and } \mathrm{HA}=\mathrm{AH}.$$

4. If A be a normal operator on a unitary space V, and B be any linear operator on V, then A is permutable with B if, and only if, A is permutable with B°.

5. Obtain an independent proof of X.7.7 by using X.7.3. [Note that any two of the three propositions IX.5.5, X.7.6 and X.7.7, imply the remaining one]

CHAPTER XI

MULTILINEAR FUNCTIONS

1. Linear functions.

Let V be a linear space over a field K. Let $f(\xi)$ be a function defined on V with values in K. The function $f(\xi)$ is said to be *linear* if, for arbitrary $\xi, \eta \in V$, and $a, b \in K$,

$$f(a\xi+b\eta)=af(\xi)+bf(\eta). \qquad \dots \quad (1)$$

Putting $a=b=0$, one obtains $f(\theta)=0$. By virtue of this property, the linear function $f(\xi)$ is said to be *homogeneous* ; and it is also called a *linear functional* or a *linear form*.

XI.1.1. *The linear functions on a linear space V, over a field K, form a linear space over K.*

Proof : Let f and g be two linear functions, and c be any scalar. Let $h=f+g$, and $k=cf$. Then, for arbitrary $\xi, \eta \in V$, and $a, b \in K$,

$$\begin{aligned} h(a\xi+b\eta) &= f(a\xi+b\eta)+g(a\xi+b\eta)=af(\xi)+bf(\eta)+ag(\xi)+bg(\eta) \\ &= ah(\xi)+bh(\eta) : \\ k(a\xi+b\eta) &= cf(a\xi+b\eta)=c[af(\xi)+bf(\eta)] \\ &= a[cf(\xi)]+b[cf(\eta)]=ak(\xi)+bk(\eta). \end{aligned}$$

Thus the sum of two linear functions and the product of a linear function by a scalar are also linear functions.

It can be readily verified that the linear functions form an abelian group, and that the requisite conditions for a linear space are all satisfied.

The linear space, formed by the linear functions on a linear space V over a field K, is called the *dual space* or the *conjugate space* of V, and will be denoted by V' (it is also denoted by Hom (V, K)).

It can readily be seen that, if $a_1, \dots, a_n$ be fixed scalars, then as ξ runs over the linear space V, of rank n,

$$f(\xi)=a_1x_1+\dots+a_nx_n$$

represents a linear function on V, where $x_1,\ldots,x_n$ are the co-ordinates of ξ, referred to a fixed basis $\alpha_1,\ldots,\alpha_n$ of V.

The converse is also true. In fact,

XI.1.2. *Referred to a basis $\alpha_1,\ldots,\alpha_n$ of the linear space V of rank n, over a field K, a linear function f on V is given by*

$$f(\xi)=a_1x_1+\ldots+a_nx_n,$$

where $a_i=f(\alpha_i)$, for $i=1,\ldots,n$, and $(x_1,\ldots,x_n)$ is the co-ordinate vector of $\xi \varepsilon V$ (referred to the above basis).

Proof: From $\xi=x_1\alpha_1+\ldots+x_n\alpha_n$, it follows, by the linearity condition, that

$$f(\xi)=f(x_1\alpha_1+\ldots+x_n\alpha_n)=x_1f(\alpha_1)+\ldots+x_nf(\alpha_n)$$
$$=a_1x_1+\ldots+a_nx_n,$$

where $a_i=f(\alpha_i)$, for $i=1,\ldots,n$.

More particularly, one obtains :

IX.1.3. *If $\alpha_1,\ldots,\alpha_n$ form a basis of a linear space V, over a field K, then for any given set of n elements, $a_1,\ldots,a_n \varepsilon K$, there exists one and only one linear function f on V, such that*

$$f(\alpha_i)=a_i, \quad \text{for } i=1,\ldots,n. \qquad \ldots \quad (2)$$

Proof: Let ξ be an arbitrary vector in V, having the co-ordinates $x_1,\ldots,x_n$, relative to the basis, $\alpha_1,\ldots,\alpha_n$. Then

$$\xi=x_1\alpha_1+\ldots+x_n\alpha_n.$$

Let a function f be defined on V, such that

$$f(\xi)=a_1x_1+\ldots+a_nx_n. \qquad \ldots. \quad (3)$$

Let η be a vector in V, having the co-ordinates $y_1,\ldots,y_n$, i.e., $\eta=y_1\alpha_1+\ldots+y_n\alpha_n$; and let c, d be any two scalars. Then

$$c\xi+d\eta=c(x_1\alpha_1+\ldots+x_n\alpha_n)+d(y_1\alpha_1+\ldots+y_n\alpha_n)$$
$$=(cx_1+dy_1)\alpha_1+\ldots+(cx_n+dy_n)\alpha_n$$

has the co-ordinates $cx_1+dy_1,\ldots,cx_n+dy_n$. Hence

$$f(c\xi+d\eta)=a_1(cx_1+dy_1)+\ldots+a_n(cx_n+dy_n)$$
$$=c(a_1x_1+\ldots+a_nx_n)+d(a_1y_1+\ldots+a_ny_n)$$
$$=cf(\xi)+df(\eta).$$

Thus the function f is linear.

Again, since the co-ordinate vectors of $\alpha_1, \alpha_2, \dots, \alpha_n$ are $(1, 0, \dots, 0)$, $(0, 1, 0, \dots, 0), \dots, (0, \dots, 0, 1)$ respectively, it follows by (3) that

$$f(\alpha_1)=a_1,\ f(\alpha_2)=a_2, \dots,\ f(\alpha_n)=a_n.$$

Next, to show that f is uniquely determined, let the contrary be assumed. So, let g be a linear function on V, such that

$$g(\alpha_i)=a_i, \text{ for } i=1, \dots, n.$$

Let $\xi=x_1\alpha_1+\dots+x_n\alpha_n$ be an arbitrary vector in V, having the co-ordinate vector $(x_1, \dots, x_n)$. Then

$$g(\xi)=g(x_1\alpha_1+\dots+x_n\alpha_n)=x_1g(\alpha_1)+\dots+x_ng(\alpha_n)$$
$$=a_1x_1+\dots+a_nx_n=f(\xi);$$

hence $g=f$.

XI.1.4. *If $\alpha_1, \dots, \alpha_n$ and $\alpha'_1, \dots, \alpha'_n$ be two bases of a linear space V, over a field K, where*

$$\alpha_i'=\Sigma p_{ij}\alpha_j, \text{ for } i=1, \dots, n,$$

and if a linear function f on V be given by

$$f(\xi)=a_1x_1+\dots+a_nx_n \text{ and } f(\xi)=a_1'x_1'+\dots a_n'x'_n,$$

with respect to the above two bases respectively, then

$$a_i'=\Sigma p_{ij}a_j, \text{ for } i=1, \dots, n.$$

[This is also expressed by saying that *the coefficients of a linear form transform cogradiently under a change of basis* (i.e., in the same way as the basis vectors)].

Proof : From the expressions for $f(\xi)$, it follows, by XI.1.2, that

$$a_i=f(\alpha_i) \text{ and } a_i'=f(\alpha_i'), \text{ for } i=1, \dots, n.$$

Thus $\quad a_i'=f(\alpha_i')=f(p_{i1}\alpha_1+\dots+p_{in}\alpha_n)$

$$=p_{i1}f(\alpha_1)+\dots+p_{in}f(\alpha_n)=p_{i1}a_1+\dots+p_{in}a_n$$

i.e., $$a_i'=\sum_{i=1}^{n} p_{ij}a_j, \text{ for } i=1 \dots, n.$$

Unitary space. Let V be, in particular, a unitary space of rank n. Then it follows immediately from the linearity property (property SP (3), page 43) that, for a fixed vector $\eta \varepsilon V$ the function f on V, defined by

$$f(\xi)=[\xi, \eta]$$

is linear. The converse is also true, as can be seen from the following :

XI.1.5. Riesz representation theorem : *If V be a finite dimensional unitary space, and f be a linear function on V, then there exists a unique vector $\eta \varepsilon V$, such that that $f(\xi)=[\xi,\eta]$, for all $\xi \varepsilon V$.*

Proof : Let $\alpha_1,\ldots,\alpha_n$ be an orthonormal basis of V. Then, by XI.1.2,

$$f(\xi)=a_1x_1+\ldots+a_nx_n,$$

where $a_1, \ldots, a_n$ are fixed scalars, and $\xi=x_1\alpha_1+\ldots.+x_n\alpha_n$.

Let η be the vector in V, defined by $\eta=a_1{}^*\alpha_1+\ldots+a_n{}^*\alpha_n$. Then

$$[\xi,\eta]=x_1a_1+\ldots.+x_na_n=a_1x_1+\ldots+a_nx_n,$$

since the basis $\alpha_1,\ldots,\alpha_n$ is orthonormal Thus there exists a vector $\eta \varepsilon V$, such that $f(\xi)=[\xi,\eta]$, for all $\xi \varepsilon V$.

To prove the uniqueness of η, let ν be also a vector, such that $f(\xi)=[\xi,\nu]$, for all $\xi \varepsilon V$. Then $[\xi,\eta]=[\xi,\nu]$, i.e., $[\xi,\eta-\nu]=0$, for all $\xi \varepsilon V$; consequently $\eta-\nu=\theta$, i.e , $\nu=\eta$.

Exercises

1. Let V be a linear space, over a field K.

(i) If S be a linear subspace of V, and $\alpha \varepsilon V-S$, then there exists a linear function f on V, such that

$$f(\alpha)=1, \text{ and } f(\xi)=0 \text{ for all } \xi \varepsilon S.$$

(ii) If α and β be two distinct vectors in V, then there exists a linear function f on V, such that $f(\alpha)\neq f(\beta)$.

(iii) If α and β be a pair of linearly independent vectors in V, then there exists a linear function f on V, such that $f(\alpha)=f(\beta)\neq 0$.

(iv) If f and g be two linear functions on V, such that $g(\xi)=0$ for a vector ξ, provided that $f(\xi)=0$, then g is a scalar multiple of f.

2. Let f be a linear function defined on a subset S of a linear space V, over a field K, such that, for scalars $a_1,\dots,a_r \varepsilon K$, and vectors $\alpha_1,\dots,\alpha_r \varepsilon S$,

$$a_1\alpha_1+\dots+a_r\alpha_r=\theta \text{ implies } a_1f(\alpha_1)+\dots+a_rf(\alpha_r)=0,$$

r being arbitrary.

Show that f can be extended to a linear function on the entire linear space V.

In particular, if the subset S be linearly independent in V, then any linear function on S can be extended to a linear function on V.

2. Dual space.

The dual space V' of a given linear space V, over a field K, is the linear space formed by the linear functions on V (with values in K). K is the field of scalars for both the linear spaces V and V'. To every pair of vectors ξ, f, where $\xi \varepsilon V$ and $f \varepsilon V'$, is associated a definite scalar $f(\xi) \varepsilon K$. The scalar $f(\xi)$ will henceforth be denoted by (ξ, f), and called the *inner product* of ξ and f. It then follows that, for any $\xi,\eta \varepsilon V$, $f, g \varepsilon V'$ and $a,b \varepsilon K$, the following properties hold ;

(a) $(\xi, af+bg)=a(\xi, f)+b(\xi, g)$.

(b) $(a\xi+b\eta, f)=a(\xi, f)+b(\eta, f)$.

The property (a) means that $(af+bg)(\xi)=af(\xi)+bg(\xi)$; and the property (b) expresses the linearity property of f, viz. $f(a\xi+b\eta)=af(\xi)+bf(\eta)$.

The inner product was previously defined for every pair of vectors belonging to the *same* unitary space (or Euclidean linear space, in particular). The concept is being generalised here. The underlying linear space V is perfectly arbitrary ; and the inner product is defined for a pair of vectors, one from V and the other one from its dual space V'.

As it is necessary to consider simultaneously vectors from both V and V', one distinguishes the vectors in V by calling them *contravariant vectors*, while those in V' are called *covariant vectors*.

A contravariant vector ξ and a covariant vector f are said to be *orthogonal*, if $(\xi, f)=0$. Two systems of contravariant vectors $\xi_1,\ldots,\xi_m$ and covariant vectors $f^1,\ldots,f^m$, containing the same number of vectors, are said to be *bi-orthogonol*, if $(\xi_i, f^j)=\delta_i^j$, where the Kronecker symbols δ_i^j have their usual meaning (i.e., $\delta_i^j=0$, if $j \neq i$, and $\delta_i^i=1$).

Dual basis. Let $\alpha_1,\ldots,\alpha_n$ from a basis of V. Let the functions f^i, for $i=1,\ldots,n$, be defined on V, by

$$f^i(\xi)=x_i, \qquad \ldots (1)$$

where $(x_1,\ldots,x_n)$ is the co-ordinate vector of ξ, with respect to the above basis. Then the functions f^i are linear ; hence they belong to the dual space V'. As the co-ordinate vector of α_j is given by $(\delta_j^1,\ldots,\delta_j^n)$, for $j=1,\ldots,n$, it follows, by (1), that

$$f^i(\alpha_j)=\delta_j^i, \text{ i.e., } (\alpha_j, f^i)=\delta_j^i ; \qquad \ldots (2)$$

hence the two systems $\alpha_1,\ldots,\alpha_n$ and $f^1,\ldots,f^n$ are bi-orthogonal.

The vectors $f^1,\ldots,f^n$ form a basis of V'. In fact, if f be an arbitary vector in V', then for a vector $\xi=x_1\alpha_1+\ldots+x_n\alpha_n \in V$,

$$\begin{aligned} f(\xi)&=x_1 f(\alpha_1)+\ldots+x_n f(\alpha_n)\\ &=a_1x_1+\ldots+a_nx_n \quad (\text{where } a_i=f(\alpha_i) \in K, \text{ for } i=1,\ldots,n)\\ &=a_1 f^1(\xi)+\ldots+a_n\ {}^{n}(\xi), \quad (\text{by } (1)) ; \end{aligned}$$

consequently f is expressible as a linear combination,

$$f=a_1f^1+\ldots+a_nf^n.$$

Also, if the vectors $f^1,\ldots,f^n$ be linearly dependent (over K), then $b_1f^1+\ldots+b_nf^n=0$, where $b_i \in K$, and $(b_1, \ldots, b_n) \neq (0,\ldots,0)$. Let $b_j \neq 0$, then $(\alpha_j, b_1f^1+\ldots+b_nf^n)=(\alpha_j, 0)=0$. But

$$(\alpha_j, b_1f^1+\ldots+b_nf^n)=b_1(\alpha_j, f^1)+\ldots+b_n(\alpha_j, f^n)=b_j.$$

Thus $b_j=0$, which is a contradiction. Hence the vectors $f^1,\ldots,f^n$ are linearly independent over K.

It follows from above that :

XI.2.1. *Corresponding to a given basis $\alpha_1,\ldots,\alpha_n$ of a linear space V, over a field K, there exists a basis $f^1,\ldots,f^n$ of the dual* *space V', such that the two systems $\alpha_1,\ldots,\alpha_n$ and $f^1,\ldots,f^n$ are bi-orthogonal.*

The basis $f^1,\ldots,f^n$ is called the *dual basis* of $\alpha_1,\ldots,\alpha_n$. The *dual basis* is uniquely determined. In fact, if $g^1,\ldots,g^n$ be also a dual basis of $\alpha_1,\ldots,\alpha_n$ then by the condition of bi-orthogonality, one obtains, as in (2), $g^i(\alpha_j)=\delta_j^i$. Then it follows, by XI.1.3, that $g^i=f^i$, for $i=1,\ldots,n$.

It follows as an immediate consequence of XI.2.1. that

Corollary: *The rank of the dual space V' is the same as the rank of V.*

It is customary to denote the contravariant vectors by letters with subscripts, while their co-ordinates are denoted by letters with superscripts. On the other hand, the covariant vectors are denoted by letters with superscripts, and their co-ordinates are denoted by letters with subscripts.

XI.2.2. *If $\alpha_1,\ldots,\alpha_n$ be a basis of a linear space V, over a field K, and $f^1,\ldots,f^n$ be the dual basis (of the dual space V'), and if a contravariant vector ξ has the co-ordinates $x^1,\ldots,x^n$, relative to the basis $\alpha_1,\ldots,\alpha_n$, and a covariant vector f has the co-ordinates $u_1,\ldots,u_n$, relative to the basis $f^1,\ldots,f^n$, then*

(i) $x^i=(\xi, f^i)$, *for* $i=1,\ldots,n$;

(ii) $u_i=(\alpha_i, f)$, *for* $i=1,\ldots,n$;

(iii) $(\xi, f)=u_1x^1+\ldots+u_nx^n$.

Proof: (i)
$$(\xi, f^i)=(x^1\alpha_1+\ldots+x^n\alpha_n, f^i)$$
$$=x^1(\alpha_1, f^i)+\ldots+x^n(\alpha_n, f^i)=x^i.$$

(ii)
$$(\alpha_i, f)=(\alpha_i, u_1f^1+\ldots+u_nf^n)$$
$$=u_1(\alpha_i, f^1)+\ldots+u_n(\alpha_i, f^n)=u_i.$$

(iii)
$$(\xi, f)=(x^1\alpha_1+\ldots x^n\alpha_n, u_1f^1+\ldots+u_nf^n)$$
$$=u_1x^1+\ldots+u_nx^n, \quad \text{since } (\alpha_i, f^j)=\delta_i^j.$$

Co-ordinate tranformation by a change of basis. By a change of basis of V, the dual basis, the co-ordinates of a contravariant and the co-ordinates of a covariant vector are transformed as follows:

XI.2.3. *If* $\alpha_1,\ldots,\alpha_n$ *and* $\alpha_1',\ldots,\alpha_n'$ *be two bases of a linear space* V, *over a field* K, *where*

$$\alpha_i'=\Sigma\, b_{ij}\alpha_j, \text{ for } i=1,\ldots,n,$$

and $f^1,\ldots,f^n$ *and* $f'^1,\ldots,f'^n$ *be their respective dual bases for the dual space* V', *and if a contravariant vector* ξ *has the co-ordinates* $x^1,\ldots,x^n$ *and* $x'^1,\ldots x'^n$, *relative to the bases* $\alpha_1,\ldots,\alpha_n$ *and* $\alpha_1',\ldots,\alpha_n'$ *respectively, and if a covariant vector* f *has the co-ordinates* $u_1,\ldots,u_n$ *and* $u_1',\ldots,u_n'$, *relative to the bases* $f^1,\ldots,f^n$ *and* $f'^1,\ldots,f'^n$ *respectively, then*

(i) $f'^i=\Sigma c_{ij}f^j$, *for* $i=1,\ldots,n$,

(ii) $x'^i=\Sigma c_{ij}x^j$, *for* $i=1,\ldots,n$,

(iii) $u_i'=\Sigma b_{ij}u_j$, *for* $i=1,\ldots,n$,

where the matrix (c_{ij}) *is the inverse of the transpose of the matrix* (b_{ij}).

(The matrix (c_{ij}) is said to be *contragradient* to the matrix (b_{ij})).

Proof : (i) Let the vectors $f^1,\ldots,f^n$ be expressed as linear combinations of the (basis) vectors $f'^1,\ldots,f'^2$, and let

$$f^j=\Sigma p_{jk}f'^k, \text{ for } j=1,\ldots,n. \qquad \ldots(3)$$

Now, $(\alpha_i'\ f^j)=(b_{i1}\alpha_1+\ldots+b_{in}\alpha_n,\ f^j)=b_{ij}$;

also $(\alpha_i',\ f^j)=(\alpha_i',\ p_{j1}f'^1+\ldots+p_{jn}f'^n)=p_{ji}$.

Hence, $p_{ji}=b_{ij}$; thus the matrix (p_{ij}) is the transpose of the matrix (b_{ij})

Now, it follows from (3) that

$$f'^i=\Sigma c_{ij}f^j, \text{ for } i=1,\ldots,n,$$

where the matrix (c_{ij}) is the inverse of the matrix (p_{ij}), so that the matrix (c_{ij}) is the inverse of the transpose of the matrix (b_{ij}).

(ii) By X.2 2(i), one obtains $x^i=(\xi,\ f^i)$ and $x'^i=(\xi,\ f'^i)$. Hence

$$\begin{aligned} x'^i &=(\xi,\ f'^i)=(\xi,\ c_{i1}f^1+\ldots+c_{in}f^n) \qquad \text{[by (i)]} \\ &=c_{i1}(\xi,\ f^1)+\ldots+c_{in}(\xi,\ f^n) \\ &=c_{i1}x^1+\ldots+c_{in}x^n=\Sigma c_{ij}x^j, \text{ for } i=1,\ldots,n. \end{aligned}$$

(iii) By XI.2.2(ii), one has $u_i=(\alpha_i, f)$ and $u_i'=(\alpha_i', f)$.

$$\text{Hence}\quad u_i'=(\alpha_i', f)=(b_{i1}\alpha_1+\dots+b_{in}\alpha_n, f)$$
$$=b_{i1}(\alpha_1, f)+\dots+b_{in}(\alpha_n, f)$$
$$=b_{i1}u_1+\dots+b_{in}u_n=\Sigma b_{ij}u_j, \text{ for } i=1,\dots,n.$$

The above results may be expressed by saying that :

A change of basis in a linear space V, over a field K,

(i) *transforms the dual basis in V' contragradiently ;*

(ii) *transforms the co-ordinates of a contravariant vector contragradiently ;*

(iii) *transforms the co-ordinates of a covariant vector cogradiently.*

Duality V of and V'. It has been shown in XI.2.2(iii) that, if ξ be a contravariant vector having co-ordinates $x^1,\dots,x^n$ relative to a basis $\alpha_1,\dots,\alpha_n$ of V, and if f be a covariant vector having co-ordinates $u_1,\dots,u_n$ relative to the dual basis $f^1,\dots,f^n$ of the dual apace V', then

$$(\xi, f)=u_1x^1+\dots+u_nx^n. \qquad \dots (4)$$

Now, if ξ be fixed, and f runs over V', then (ξ, f) is a linear function on V', given by (4).

Conversely, any linear function on V' is, by XI.1.2, expressible as $u_1x^1+\dots+u_nx^n$, where $x^1,\dots,x^n$ are fixed scalars, and $u_1,\dots,u_n$ are the co-ordinates of the current vector f, referred to a basis $f^1,\dots,f^n$ of V'. If ξ be the (fixed) vector in V, whose co-ordinates are $x^1,\dots,x^n$, relative to the basis $\alpha_1,\dots,\alpha_n$ of V, bi-orthogonal to the basis $f^1,\dots,f^n$ of V', it follows that $u_1x^1+\dots+u_nx^n=(\xi, f)$. Hence every linear function on V' determines uniquely a vector $\xi \in V$, such that (ξ, f) represents the given linear function.

Thus there exists a one-to-one correspondence between the linear functions on V' and the vectors in V. The vectors in V can therefore be considered as linear functions on V'. In other words, *V can be considered as the dual space of V'.*

Dual of a linear operator. Let $\mathbf{A}$ be a linear operator on a linear space V, over a field K. Let $\mathbf{A}'$ be an operator on the dual space V', defined by

$$(\xi\mathbf{A}, f)=(\xi, f\mathbf{A}'), \text{ where } \xi \varepsilon V, f \varepsilon V'. \qquad \ldots (5)$$

The operator $\mathbf{A}'$ is linear. In fact,

$$\begin{aligned}(\xi, (af+bg)\mathbf{A}')&=(\xi\mathbf{A}, af+bg)=a(\xi\mathbf{A}, f)+b(\xi\mathbf{A}, g)\\ &= a (\xi, f\mathbf{A}')+b(\xi, g\mathbf{A}'),\end{aligned}$$

where $a, b \varepsilon K$, $f, g \varepsilon V'$ and $\xi \varepsilon V$.

The linear operator $\mathbf{A}'$, as defined above, is called the *conjugate* or the *dual of the linear operator* $\mathbf{A}$.

It follows readily from the definition that, if $\mathbf{A}$ and $\mathbf{B}$ be two linear operators on V, and $a, b \varepsilon K$, then

(i) $(a\mathbf{A}+b\mathbf{B})'=a\mathbf{A}'+b\mathbf{B}'$, and (ii) $(\mathbf{A}\mathbf{B})'=\mathbf{B}'\mathbf{A}'$.

The proof is left as an exercise.

Let $\mathbf{A}$ be represented by the matrix $A=(a_{ij})$, relative to a basis $\alpha_1,\ldots,\alpha_n$ of V, and let $\mathbf{A}'$ be represented by the matrix $A'=(a_{ij}')$, with respect to the dual basis $f^1,\ldots,f^n$ of the dual space V'. Then

$$(\alpha_i, f^j\mathbf{A}')=(\alpha_i, a_{j1}'f^1+\ldots\ldots a_{jn}'f^n)=a_{ji}',$$

also $$(\alpha_i, f^j\mathbf{A}')=(\alpha_i\mathbf{A}, f^j)=(a_{i1}\alpha_1+\ldots+a_{in}\alpha_n, f^j)=a_{ij}.$$

Hence $a_{ji}'=a_{ij}$; in other words, the matrix A' is the transpose of the matrix A. Thus one obtains :

XI.2.4. *If a linear operator* $\mathbf{A}$ *on a linear space* V, *over a field* K, *be represented by the matrix* A, *relative to basis* $\alpha_1,\ldots,\alpha_n$ *of* V, *then the dual linear operator* $\mathbf{A}'$ *is represented by* A^t, *the transpose of* A, *when referred to the dual basis* $f^1,\ldots,f^n$ *of the dual space* V'.

Unitary space. Let V be, in particular, the unitary space of rank n. It follows from Riesz representation theorem (XI.1.5) that for any linear function f on V, there exists a unique vector $\eta \varepsilon V$, such that $(\xi, f)=[\xi, \eta]$, where ξ is an arbitrary vector in V ; and conversely. Thus, there exists a one-to-one correspondence between the vectors in V' with the vectors in V. The dual space V' can therefore be identified with the space V. This is sometimes expressed by saying that *in the case of a unitary space, one can replace the covariant vectors by contravariant vectors.*

Let $\alpha_1, \ldots, \alpha_n$ be a basis of V, and let $f^1, \ldots, f^n$ be the dual basis in V'. When V' is identified with V, $f^1, \ldots, f^n$ are to be considered as vectors belonging to V; and they form a basis of V. Let

$$\alpha_i = \Sigma g_{ij} f^j, \text{ for } i=1,\ldots,n. \qquad \ldots \quad (6)$$

Then $$[\alpha_i, \alpha_j] = [g_{i1} f^1 + \ldots + g_{in} f^n, \alpha_j] = g_{ij};$$

thus $$g_{ij} = [\alpha_i, \alpha_f], \text{ for } i, j = 1, \ldots, n.$$

Now, solving the equations (6) for f^i, one obtains

$$f^i = \Sigma g^{ij} \alpha_j, \text{ for } i = 1, \ldots, n,$$

where the matrix (g^{ij}) is the inverse of the matrix (g_{ij}), i.e.,

$$\Sigma g^{ik} g_{kj} = \delta_j^i, \text{ for } i, j = 1, \ldots . n.$$

Exercises

1. Let V be a finite dimensional linear space, over a field K. Let S be a *subset* of V, and A(S) be the subset of the dual space V', consisting of all those linear functions f on V, such that $(\xi, f) = 0$ for all $\xi \in S$. Then A(S) is called the *annihilator of* S. Prove that :

(i) A(S) is a linear subspace of V';

(ii) A(A(S)) coincides with the linear subspace spanned by S;

(iii) if S and T be two subsets of V, then $S \subset T$ implies A(T) $\subset$ A(S);

(iv) if S be a linear subspace of V, then

(a) A(A(S)) $= S$;

(b) rank A(S) $=$ rank $V -$ rank S;

(v) if S and T be two linear subspaces of V, then

(a) A($S \cap T$) $=$ A(S) $+$ A(T),

(b) A($S + T$) $=$ A(S) $\cap$ A(T).

2. Show that exercise 1 (iv-b) can be applied to prove V.1.2 for a system of linear homogeneous equations.

3 Multilinear functions.

Let V be a linear space, over a field K, and let V' be the dual space. Then a function $F(\xi, \eta, \ldots, f, g, \ldots)$, with values in K, is called a *multilinear function* of p contravariant vectors $\xi, \eta, \ldots$ and q covariant vectors $f, g, \ldots$, if F is linear in each of its arguments. It is also called a multilinear function of *type* (p, q). The total number of arguments of a multilinear function is called its *valence*. Thus the function F has the valence $p+q$.

Linearity in the first contravariant argument means that

$$F(a\xi+b\xi', \eta, \ldots, f, g, \ldots)$$
$$=aF(\xi, \eta, \ldots, f, g, \ldots)+bF(\xi', \eta, \ldots, f, g, \ldots);$$

likewise, linearity in the first covariant argument means

$$F(\xi, \eta, \ldots, af+bf', g, \ldots)$$
$$=aF(\xi, \eta, \ldots, f, g, \ldots)+bF(\xi, \eta, \ldots, f', g, \ldots);$$

a, b being arbitrary scalars.

Multilinear functions of valence 1 are of the two types $(1,0)$ and $(0,1)$. A multilinear function of type $(1,0)$ is a linear function of one contravariant vector, and is therefore a covariant vector. On the other hand, a multilinear function of type $(0,1)$ is a linear function of one covariant vector; and by the duality between V and V', it is a contravariant vector.

Multilinear functions of valence 2 are of the three types $(2,0)$, $(1,1)$ and $(0,2)$. A multilinear function of type $(2,0)$ is a bilinear function on V, while a multilinear function of type $(0,2)$ is a bilinear function on V'. A multilinear function of type $(1,1)$ is called a *mixed bilinear function.*

Let $F(\xi, f)$ be a mixed bilinear function. Then, for a fixed ξ, $F(\xi, f)$ is a linear function on V'; hence there corresponds to $F(\xi, f)$ a unique vector $\eta \varepsilon V$, such that $F(\xi, f)=(\eta, f)$. Let $\mathbf{A}$ be an operator on V, defined by

$$\xi\mathbf{A}=\eta, \qquad \text{i.e.,} \quad F(\xi, f)=(\xi\mathbf{A}, f). \qquad \ldots (1)$$

By the linearity of F in its first argument, it follows readily that the operator $\mathbf{A}$ is linear.

Again, corresponding to a given linear operator A on V, there exists a mixed bilinear function F, given by (1). It can readily be shown that the correspondence between the mixed bilinear functions F and the linear operators A on V, is one-to-one. Also

$$F(\xi, f)=(\xi \mathrm{A}, f) \text{ and } G(\xi, f)=(\xi \mathrm{B}, f)$$

imply $aF(\xi, f)+bG(\xi, f)=(\xi(a\mathrm{A}+b\mathrm{B}), f)$, for any $a,b \in K$.

Thus, if the linear operators A and B correspond respectively to the mixed bilinear functions F and G, then $a\mathrm{A}+b\mathrm{B}$ corresponds to $aF+bG$. This is sometimes expressed by saying that the one-to-one correspondence between the mixed bilinear functions and the linear operators on V is linear.

Sum and product. The sum of two multilinear functions is defined, provided that they are of the same type. Let $F_1(\xi,\eta,\ldots,f,g,\ldots)$ and $F_2(\xi,\eta,\ldots,f,g,\ldots)$ be two multilinear functions of the same type (p,q) on the same linear space V. Then their sum is defined by

$$F_1(\xi,\eta,\ldots,f,g,\ldots,\ldots)+F_2(\xi,\eta,\ldots,f,g,\ldots)=F(\xi,\eta,\ldots,f,g,\ldots).$$

It can readily be verified that the sum F is also a multilinear function of the type (p, q).

The product of a multilinear function by a scalar is evidently a multilinear function.

The product can, however, be defined for any two multilinear functions on the same linear space V. Let $F_1(\xi,\eta\ldots,f,g,\ldots)$ and $F_2(\zeta,\chi,\ldots,h,k,\ldots)$ be two multilinear functions of types (p_1,q_1) and (p_2,q_2) respectively. Then their product F is defined to be a multilinear function F of type (p_1+p_2, q_1+q_2), given by

$$F_1(\xi,\eta,\ldots,f,g,\ldots).F_2(\zeta,\chi,\ldots,h,k,\ldots)=F(\xi,\eta,\ldots,\zeta,\chi,\ldots,f,g,\ldots,h,k,\ldots).$$

In forming the sum of two multilinear functions (of the same type), the corresponding arguments are identified. While, in forming the product of two multilinear functions, the arguments in the two functions are treated as independent, and are arranged in a prescribed manner.

Coordinates relative to a basis Let $\alpha_1,\ldots,\alpha_n$ be a basis of V, and $f^1,\ldots,f^n$ the dual basis in V'. Let

$$F(\alpha_i,\alpha_j,\ldots,f^r,f^s,)\ldots) \text{ be denoted by } F^{rs\cdots}_{ij\cdots} \qquad \ldots. (2)$$

If, now, the contravariant vectors $\xi,\eta,\ldots$ have the co-ordinates $(x^1,\ldots,x^n)$, $(y^1,\ldots,y^n)$, .. respectively, relative to the basis $\alpha_1,\ldots,\alpha_n$, i.e., $\xi=x^1\alpha_1+\ldots+x^n\alpha_n$, $\eta=y^1\alpha_1+\ldots+y^n\alpha_n,\ldots,$ and the covariant vectors $f,g,\ldots$ have the co-ordinates $(\phi_1,\ldots,\phi_n)$, $(\psi_1,\ldots,\psi_n),\ldots$ respectively, relative to the basis $f^1,\ldots,f^n$, i.e., $f=\phi_1 f^1+\ldots+\phi_n f^n$, $g=\psi_1 f^1+\cdots+\psi_n f^n,\ldots,$ then it follows, by the property of linearity of F in each of its arguments, that

$$F(\xi,\eta,\ldots,f,g,\ldots)=F(\Sigma x^u\alpha_u,\Sigma y^v\alpha_v,\ldots,\Sigma\phi_r f^r,\Sigma\psi_s f^s,\ldots)$$

$$=\Sigma F(\alpha_u,\alpha_v,\ldots,f^r,f^s,\ldots)x^u y^v\ldots\phi_r\psi_s\ldots$$

$$=\Sigma F^{rs\cdots}_{uv\cdots}\, x^u y^v\ldots\phi_r\psi_s\ldots. \qquad \ldots. (3)$$

The numbers $F^{rs\cdots}_{uv\cdots}$, which determine the multilinear function F, are called the *co-ordinates of the multilinear function* F, *relative to the basis* $\alpha_1,\ldots,\alpha_n$ of V.

Co-ordinate transformation. A change of basis induces a transformation of co-ordinates of a multilinear function in the following way.

XI.3.1. *If* $\alpha_1,\ldots,\alpha_n$ *and* $\alpha_1',\ldots,\alpha_n'$ *be two bases of a linear space* V, *over a field* K, *where* $\alpha_i'=\Sigma b_i^j\alpha_j$, *for* $i=1,\ldots,n$, *and* $f^1,\ldots,f^n$ *and* $f'^1,\ldots,f'^n$ *be their respective dual bases in the dual space* V', *and if a multilinear function* $F(\xi,\eta,\ldots,f,g,\ldots)$ *have the co-ordinates* $F^{km\cdots}_{uv\cdots}$ *with respect to the bases* $\alpha_1,\ldots,\alpha_n$ *and* $f^1,\ldots,f^n$, *and have the co-ordinates* $F'^{rs\cdots}_{ij\cdots}$ *relative to the bases* $\alpha_1',\ldots,\alpha_n'$ *and* $f'^1,\ldots,f'^n$, *then*

$$F'^{rs\cdots}_{ij\cdots}=\Sigma b_i^u b_j^v\ldots c_k^r c_m^s\, F^{km\cdots}_{uv\cdots}, \qquad \ldots (4)$$

where the matrix (c_i^j) *is the transpose of the inverse of the matrix* (b_i^j).

Proof: Let $f'^j = \Sigma c_i^j f^i$, for $i=1,\ldots,n$; then the matrix (c_i^j) is the transpose of the inverse of the matrix (b_i^j), by XI.2.3(i).

For a fixed i, the co-ordinates of α_i', relative to the basis $\alpha_1, \ldots, \alpha_n$, are given by $(b_i^1, \ldots, b_i^n)$; and for a fixed j, the co-ordinates of f'^j, relative to the basis $f^1, \ldots, f^n$, are given by $(c_1^j, \ldots, c_n^j)$. It now follows from (2) and (3) that

$$\begin{aligned} F'^{rs\ldots}_{ij\ldots} &= F(\alpha_i', \alpha_j', \ldots, f'^r, f'^s, \ldots) \\ &= \Sigma b_i^u b_j^v \ldots c_k^r c_m^s \ldots F^{km\ldots}_{uv\ldots}. \end{aligned}$$

The formula (4) may be described briefly by saying that *the matrix* (b_i^j) *operates on the lower indices of* $F^{km\ldots}_{uv\ldots}$, *while the upper indices are affected by the matrix* (c_i^j).

Exercises

1. Show that the sum and the product of two multilinear functions, as defined above, are also multilinear functions.

2. Find whether the multilinear functions of the type (p, q), over a linear space V of rank n, over a field, K, form a linear space; and, if so find its rank.

4. Tensors.

Let V be a linear space of rank n, over a field K. It has been shown that:

(a) *with respect to a given basis of* V,

(i) a vector belonging to V is represented by its n co-ordinates, which are elements of K;

(ii) a linear function on V is determined by the n coefficients of the corresponding linear form, and these coefficients are elements of K;

(iii) a linear operator on V is represented by an $n \times n$ matrix over K, i.e., by n^2 elements of K;

(iv) a multilinear fuction of type (p, q) is determined by its n^{p+a} components $F^{uv\cdots}_{rs\cdots}$, which are elements of K;

and, (b) *by a change from a basis* $\alpha_1,\ldots,\alpha_n$ to $\alpha_1',\ldots,\alpha_n'$, *where* $\alpha_i'=\Sigma b_i^j\alpha_j$, *for* $i=1,\ldots,n$,

(i) the co-ordinates of a vector ξ are transformed from $(x_1, \ldots,x_n)$ to $(x_1',\ldots,x_n')$, where $x_i=\Sigma b'^j_i x_j$, for $i=1,\ldots,n$, the matrix $(b_i'^j)$ being the transpose of the matrix (b_i^j);

(ii) the coefficients of the linear form, which represents a given linear function, are transformed from $(a_1,\ldots,a_n)$ to $(a_1',\ldots, a_n')$, where $a_i'=\Sigma b_i^j a_j$, for $i=1,\ldots,n$;

(iii) the matrix representing a linear operator on V is transformed from A to A', where $A'=BAB^{-1}$, with $B=(b_{ij})$;

(iv) the components of a given multilinear function are transformed from $F^{km\cdots}_{uv\cdots}$ to $F'^{rs\cdots}_{ij\cdots}$, where

$$F'^{rs\cdots}_{ij\cdots}=\Sigma b_i^u b_j^v\ldots c_k^r c_m^s\ldots F^{km\cdots}_{uv\cdots},$$

the matrix (c_i^j) being the inverse of the transpose of the matrix (b_i^j).

Each of the above objects is represented by a system of scalars (called its co-ordinates, components, etc) with respect to a given basis of V, and these scalars are transformed in a particular manner, under a change of the basis. Conversely, each of these objects is completely determined when its associated scalars (i.e., its co-ordinates, components, etc), relative some basis of V, are given, and the law of transformation of these scalars, under a change of basis V is prescribed.

A closely related concept is now introduced :

A p times covariant and q times contravariant tensor is said to be defined over an n-dimensional linear space V, over a field K, if

(a) with every basis of V, there is associated a set of n^{p+a} scalars $F^{uv\cdots}_{rs\cdots}$ belonging to K, where there are p lower indices and q upper indices, and

(b) by a change from a basis $\alpha_1,\ldots,\alpha_n$ to $\alpha_1',\ldots,\alpha_n'$ of V, where $\alpha_i'=\Sigma b_i^j\alpha_j$, for $i=1,\ldots,n$, the associated scalars

$$F_{uv\cdots}^{km\cdots} \text{ are transformed to } F'^{rs\cdots}_{ij\cdots},$$

according to the rule $$F'^{rs\cdots}_{ij\cdots}=\Sigma b_i^u b_j^v\ldots c_k^r c_m^s\ldots F_{uv\cdots}^{km\cdots},$$

where the matrix (c_i^j) is the inverse of the transpose of the matrix (b_i^j).

The scalars $F_{rs\cdots}^{uv\cdots}$ are called the *components of the tensor*, relative to the basis $\alpha_1,\ldots,\alpha_n$ of V. The number $p+q$ is called the *valence* or the *rank of the tensor*.

It follows from the above definition that a multilinear function of type (p, q) determines uniquely a tensor, which is p times covariant and q times contravarint ; and conversely, a p-times covariant and q times contravariant tensor determines uniquely a multilinear function of type (p,q). Thus multilinear functions provide with one of the possible realisations of tensors. And, properties of tensors can be deduced from those of multilinear functions.

It now follows immediately,

(i) by a(i) and b(i), that any contravariant vector, i.e., a vector belonging to V is a contravariant tensor of rank 1 ;

(ii) by a(ii) and b(ii), that any linear function on V, i.e., a covariant vector is a covariant tensor of rank 1 ;

(iii) by a(iii) and b(iii), that any linear operator on V is a mixed tensor of rank 2, covariant in the first index (or the lower index) and contravariant in the second index (or the upper index).

As two tensors of the same type are transformed in the same way, it follows that *two tensors of the same type are equal, if their corresponding components, relative to some basis of V, are equal* (in this case their corresponding components are equal relative to every basis of V).

Sum and product. One defines sum and product of two tensors in the same way as in the case of multilinear functions.

Given two tensors of the same type, with components $F'^{rs\cdots}_{ij\cdots}$ and $F''^{km\cdots}_{uv\cdots}$, their sum is defined to be the tensor, whose components are obtained by adding the corresponding components of the two given tensors ; thus the components of the sum are given by

$$F^{rs\cdots}_{ij\cdots} = F'^{rs\cdots}_{ij\cdots} + F''^{rs\cdots}_{ij\cdots}.$$

Given two tensors, not necessarily of the same type, with components $F'^{rs\cdots}_{ij\cdots}$ and $F''^{km\cdots}_{uv\cdots}$, their product is defined to be the tensor, whose components are given by

$$F^{rs\cdots km\cdots}_{ij\cdots uv\cdots} = F'^{rs\cdots}_{ij\cdots} F''^{km\cdots}_{uv\cdots}.$$

If the first tensor be p_1 times covariant and q_1 times contravariant, and the second be p_2 times covariant and q_2 times contravariant, then their product is a (p_1+p_2) times covariant and (q_1+q_2) times contravariant tensor.

Contraction. Contraction is a method by which, from a tensor which is p times covariant and q times contravariant, one obtains a tensor which is $(p\text{-}1)$ times covariant and $(q\text{-}1)$ times contravariant ; the rank of the tensor is thus depressed by 2.

Let a p times covariant and q times contravariant tensor be given, whose components, with respect to a basis $\alpha_1, \ldots, \alpha_n$ of the underlying linear space V (over a field K), be $F^{rs\cdots}_{ij\cdots}$. Let $F(\xi, \eta, \ldots f, g, \ldots)$ be the corresponding multilinear function in p contravariant vectors and q covariant vectors. Then

$$F^{rs\cdots}_{ij\cdots} = F(\alpha_i, \alpha_j, \ldots, f^r, f^s, \ldots),$$

where $f^1, \ldots, f^n$ is the dual basis in dual space V'. Now, a function F' is defined as follows :

$$F'(\eta, \ldots, g, \ldots) = F(\alpha_1, \eta, \ldots, f^1, g, \ldots) + F(\alpha_2, \eta, \ldots f^2, g, \ldots) + \ldots$$
$$+ F(\alpha_n, \eta, \ldots, f^n, g, \ldots). \quad \ldots (1)$$

As each summand is linear in respect of each of the p-1 contravariant vectors $\eta, \ldots,$ and also in respect of each of the

q-1 covariant vectors $g, \ldots$, it follows that the function F' is multilinear in (p-1) contravariant vectors $\eta, \ldots$, and q-1 covariant vectors $g, \ldots$.

It will now be shown that, although the summands in (1) depend on the particular bases $\alpha_1, \ldots, \alpha_n$ and $f^1, \ldots, f^n$, the function F' is independent of the choice of bases for V and V'. Let $\alpha_1', \ldots, \alpha_n'$ be another basis of V, and $f'^1, \ldots, f'^n$ be the dual basis in V'.

Since the contravariant vectors $\eta, \ldots$, and the covariant vectors $g, \ldots$, remain fixed, under the change of bases, $F(\alpha_i, \eta, \ldots, f^i, g, \ldots)$ may be considered as a bilinear function $G(\alpha_i, f^i)$, for $i=1, \ldots, n$. It then suffices to show that

$$G(\alpha_1, f^1) + \ldots + G(\alpha_n, f^n) = G(\alpha_1', f'^1) + \ldots + G(\alpha_n', f'^n). \qquad \ldots (2)$$

Let $\alpha_i' = \Sigma b_i^j \alpha_j$, for $i = 1 \ldots, n$, then $f'^j = \Sigma b_i^j f^i$, for $j = 1, \cdots, n$.

Now,
$$\begin{aligned} G(\alpha_i', f'^i) &= G(b_i^1 \alpha_1 + \ \ + b_i^n \alpha_n,\ b_1^i f^1 + \ldots + b_n^i f^n) \\ &= G(\alpha_i, f^i), \qquad \text{for } i = 1, \ldots, n. \end{aligned}$$

Hence (2) holds; consequently, the multilinear function F', as defined in (1), is independent of the choice of bases for V and V'. Let

$F'(\alpha_j, \ldots, f^s)$ be denoted by ${F'}^{s\cdots}_{j\cdots}$.

Also,

$$F'(\alpha_j, \ldots f^s) = F(f_1, \alpha_j \ldots, f^1, f^s, \ldots) + \ldots + F(\alpha_n, \alpha_j, \ldots, f^n, f^s, \ldots).$$

Hence, it follows that

$${F'}^{s\cdots}_{j\cdots} = F^{1s\cdots}_{1j\cdots} + F^{2s\cdots}_{2j\cdots} + \ldots + F^{ns\cdots}_{nj\cdots}. \qquad \ldots (3)$$

Thus a (p-1) times covariant and (q-1) times contravariant tensor has been obtained, whose components, referred to the basis $\alpha_1, \ldots, \alpha_n$, are given by ${F'}^{s\cdots}_{j\cdots}$ (which are determined in terms of the components of the given tensor by (3)).

Euclidean space. Let V be an n-dimensional Euclidean space. It has been shown (in general, for a unitary space) that there exists an isomorphism between V and its dual space V', such that, for any vector $f \varepsilon V'$, there exists a unique vector $\zeta \varepsilon V$, satisfying the relation $(\xi, f) = [\xi, \zeta]$, for all $\xi \varepsilon V$. Hence given a multilinear function $F(\xi,\eta,\ldots,f,g,\ldots)$ in p contravariant vectors $\xi,\eta,\ldots,$ and q covariant vectors $f, g,\ldots,$ one obtains, correspondingly, a function $F(\xi,\eta,\ldots,\zeta,\chi,\ldots)$ in $p+q$ contravariant vectors $\xi,\eta,\ldots,\zeta,\chi,\ldots,$ by replacing the covariant vectors $f,g,\ldots,$by the contravariant vectors $\zeta,\chi,\ldots.$

Let $\alpha_1,\ldots,\alpha_n$ be a basis of V. If $F^{rs\cdots}_{ij\cdots}$ be the co-ordinates of the multilinear function $F(\xi,\eta,\ldots,f,g,\ldots)$, relative to the above basis, then

$$F^{rs\cdots}_{ij\cdots}=F(\alpha_i,\alpha_j,\ldots f^r,f^s,\ldots),$$

where $f^1,\ldots,f^n$ form the dual basis in V'.

The co-ordinates of the multilinear function $F(\xi,\eta,\ldots,\zeta\ \chi,\ldots)$, relative to the above bases, are given by

$$F_{ij\cdots rs\cdots}=F(\alpha_i,\alpha_j,\ldots,\alpha_r\ \alpha_s,\ldots).$$

It has been shown that the vectors of the basis $\alpha_1,\ldots,\alpha_n$ are given, in terms of the vectors $f^1,\ldots,f^n$ in the dual basis, by

$$\alpha_r=\Sigma g_{ku}f^u,\qquad \text{where } g_{ku}=[\alpha_k,\alpha_u].$$

Hence, one obtains

$$\begin{aligned} F_{ij\cdots rs\cdots}&=F(\alpha_i,\alpha_j,\ldots,\alpha_r,\alpha_s,\ldots)\\ &=F(\alpha_i,\alpha_j,\ldots,\Sigma g_{ru}f^u,\Sigma g_{sv}f^v,\ldots)\\ &=\Sigma g_{ru}g_{sv}\cdots F(\alpha_i,\alpha_j,\ldots,f^u,f^v,\ldots)\end{aligned}$$

i.e.,

$$F_{ij\cdots rs\cdots}=\Sigma g_{ru}g_{sv}\cdots F^{uv\cdots}_{ij\cdots}. \qquad \ldots (4)$$

One obtains, correspondingly, the following result for tensors :

If $F^{rs\cdots}_{ij\cdots}$ *be the components of a tensor in a Euclidean space, which is* p *times covariant and* q *times contravariant, then one can construct a new tensor, which is* $p+q$ *times covariant, with components* $F_{ij\cdots rs\cdots}$ *given by* (4), *where* $g_{ik}=[\alpha_i,\alpha_k]$.

This operation of obtaining a $p+q$ times covariant tensor from a given tensor, which is p times covariant and q times contravariant, is known as *lowering of indices of a tensor.*

Similarly, *given a* p *times covariant and* q *times contravariant tensor, with components* $F^{rs\cdots}_{ij\cdots}$, *in a Euclidean space, one can* *obtain a new tensor, which is* $p+q$ *times contravariant, whose components* $F^{ij\cdots rs\cdots}$ *are given by*

$$F^{ij\cdots rs\cdots}=\Sigma g^{ui}g^{vj\cdots}F^{rs\cdots}_{uv\cdots}, \text{ where } g^{ui}=[f^u,f^i].$$

This operation is known as *raising of indices of a tensor.*

The operation of lowering of indices of a tensor may also be considered as contraction of the product of the given tensor and some twice covariant metric tensors g_{ij} (see exercise 5). Likewise, the operation of raising of indices of a tensor may be effected by contraction of a product of the given tensor by some twice contravariant metric tensors g^{ik}.

Contraction of a mixed tensor of rank 2 gives a scalar. Contraction of a (mixed) tensor of rank 3 gives a tensor of rank 1, i.e., a vector (covariant or contravariant). Thus carrying on the operation of contraction, over and over again, one obtains tensors of lower rank (till all the upper or lower indices are exhausted).

On the other hand, one can obtain tensors of higher rank, by multiplying tensors of rank 1, i.e., vectors (covariant and contravariant) (vide exercise 2). It should however be noted that it is not generally possible to obtain any given tensor by multiplying vectors. But, it can be shown that every tensor can be obtained from suitable vectors (covariant and contravariant) by the operations of addition and multiplication.

Exercises

1. Show that the components of the tensor, corresponding to the identity transformation on a linear space V, are given by δ_i^j, with respect to any basis of V, whatsoever.

2. Show that, if $x^1,\ldots,x^n$ be the co-ordinates of a contravariant vector (referred to a basis of V) and $u_1,\ldots,u_n$ are the co-ordinates of a covariant vector (referred to the dual basis in V'), then $g_j^i=x^iu_j$ are the components of a mixed tensor of rank 2.

3. Show that, given p and q, it is always possible to construct a p times covariant and q times contravariant tensor,

over a linear space V of rank n, whose components assume n^{p+q} prescribed values, with respect to some basis of V.

4. Let V be a linear space of rank n, over a field K; and let V' be the dual space.

A function $F(\xi\,;\mu)$ is said to be a *bilinear function* or *bilinear form* in the vectors ξ and μ, if

(a) for any fixed μ, $F(\xi\,;\mu)$ is a linear function of ξ, and

(b) for any fixed ξ, $F(\xi\,;\mu)$ is a linear function of μ.

Three cases may arise :

(i) Let $\xi, \mu \varepsilon V$; then referred to a basis of V, $F(\xi\,;\mu)$ is represented by a matrix A of order n, over K.

(ii) Let $\xi \varepsilon V$ and $\mu \varepsilon V'$; then referred to a basis of V and the dual basis in V', $F(\xi\,;\mu)$ is represented by a matrix A of order n, over K.

(iii) Let $\xi, \mu \varepsilon V'$; then referred to a basis of V', $F(\xi\,;\mu)$ is represented by a matrix A of order n, over K.

Associating the matrix A with the corresponding basis, one obtain a tensor of rank 2 in all the three cases.

Show that the tensor is twice covariant in case (i), twice contravariant in case (iii), and once covariant and once contravariant in case (ii).

5. Show that (i) g_{ik} are the components of a twice covariant tensor, and (ii) g^{ik} are the components of a twice contravariant tensor, where g_{ik} and g^{ik} are as defined above in Art. 4. [*Hints* : $g_{ik}=[\alpha_i, \alpha_k]$ are the coefficients of a bilinear form, viz. the inner product relative to the basis $\alpha_1, \ldots, \alpha_n$.]

(The tensors g_{ik} and g^{ik} are known as *metric tensors*).

4.1. Symmetric and skew-symmetric tensors.

Let $F(\xi, \eta, \ldots, f, g, \ldots)$ be a multilinear function of type (p, q) over a linear space V (over a field K). It is evident that, by any permutation of a set argument vectors, all of which are contravariant or all covariant, F is tranformed to a multilinear function. The multilinear function F is said to be *symmetric with respect to a set of argument vectors*, all of which are contravariant or all covariant, if it remains invariant under every permutation of those argument vectors. Accordingly, a tensor is said to be *symmetric, with respect to a given set of indices*, all of which are upper or all lower, if its components remain invariant under every permutation of these indices.

A multilinear function $F(\xi, \eta, \ldots)$, all whose arguments are contravariant vectors or all covariant vectors, is said to be *skew-symmetric* or *anti-symmetric in a prescribed set of argument vectors*, if the function is transformed to its negative, by interchange of any two of these arguments. Accordingly, a tensor, all whose indices are upper or all lower, is said to be *skew-symmetric* or *anti-symmetric in a prescribed set of indices*, if it changes its sign, whenever two of these indices are interchanged. The skew-symmetric tensor changes its sign by every odd permutation of the prescribed indices, while it remains unaltered by every even permutation of the prescribed indices.

Since, for a skew-symmetric tensor of rank n, components which differ only in the arrangement of the indices are either equal or negative of each other, it follows that the number of independent non-null components in r indices is ${}^nC_r$. If $r=n$, the non-null components are all equal in magnitude and may differ in sign only. If r exceeds n, then the (skew-symmetric) tensor has no non-null component.

Symmetrization. Given a tensor, one can construct another tensor which is symmetric in a prescribed set of indices (all of which are lower or all upper) in the following way. This process is known as *symmetrization*.

Let Γ be a given tensor, and let $i_1, \ldots, i_s$ be a prescribed set of upper indices, occuring in Γ. Then

$$\Lambda = \frac{1}{s\,!}\,\Sigma\sigma(\Gamma),$$

where the summation extends over all permutations σ of the indices $i_1, \ldots, i_s$ (and where $\sigma(\Gamma)$ denotes the tensor which is obtained by applying the permutation σ on the upper indices of

Γ), is evidently a tensor, symmetric in the prescribed set of indices $i_1,\ldots,i_s$.

The operation of symmetrization can be carried out with respect to a prescribed set of upper indices, or with respect to a presecribed set of lower indices ; but not with respect to a set of mixed indices. The operation of symmetrization may be indicated by enclosing the prescribed set of indices in parenthesis. For example,

$$\Gamma^{(ij)k}=\tfrac{1}{2}(\Gamma^{ijk}+\Gamma^{jik}),$$

$$\Gamma^{i(jkm)}=\tfrac{1}{6}(\Gamma^{ijkm}+\Gamma^{ijmk}+\Gamma^{imkj}+\Gamma^{imjk}+\Gamma^{ikjm}+\Gamma^{ikmj}).$$

Alternation. Given a tensor, one can construct another tensor which is skew-symmetric in a prescribed set of indices (all of which are lower or all upper) in the following way. This operation is known as *alternation* or *alternatization.*

Let Γ be a given tensor, and let $i_1,\ldots,i_s$ be a prescribed set of upper indices, occuring in Γ, Then

$$\Lambda=\frac{1}{s\,!}\Sigma\pm\sigma(\Gamma),$$

where the summation extends over all permutations σ of the indices $i_1,\ldots,i_s$, the positive sign being taken for even permutations and the negative sign for odd permutations, is a tensor which is skew-symmetric in the prescribed set of indices $i_1,\ldots,i_s$. The operation of alternation is usually indicated by enclosing the prescribed set of indices within a square bracket. For example,

$$\Gamma^{[ij]k}=\tfrac{1}{3}(\Gamma^{ijk}-\Gamma^{jik}),$$

$$\Gamma^{i[jkm]}=\tfrac{1}{6}(\Gamma^{ijkm}-\Gamma^{ikjm}+\Gamma^{ikmj}-\Gamma^{imkj}+\Gamma^{imjk}-\Gamma^{ijmk}).$$

Let $\xi,\eta,\ldots,\zeta$ be k given vectors in V, and let F be the tensor product of these vectors (the vectors being considered as tensors of rank 1). By the operation of alternation, one obtains from F a skew-symmetric tensor G. Now, referred to a basis $\alpha_1,\ldots,\alpha_n$ of V, the vectors have their respective co-ordinates $(x^1,\ldots,x^n),(y^1,\ldots,y^n),\ldots,(z^1,\ldots,z^n)$. It can readily be shown that the components of G relative to the above basis of V, are given by all the kth order minors of the following matrix.

$$\begin{pmatrix} x^1 & x^2 & \dots & x^n \\ y^1 & y^2 & \dots & y^n \\ \dots & \dots & \dots & \dots \\ z^1 & z^2 & \dots & z^n \end{pmatrix}. \qquad \dots (1)$$

The tensor G remains unaltered, when any of the vectors $\xi, \eta, \dots, \zeta$, is replaced by the sum of that vector and any linear combination of the remaining vectors.

Let S be a linear subspace of rank k of the linear space V; and let, $\xi, \eta, \dots, \zeta$ form a basis of S. Then one can associate S with the skew-symmetric tensor G, constructed with the vectors $\xi, \eta, \dots, \zeta$, as described above. The components of G are called the *Plucker's co-ordinates* of the linear subspace S.

It can be shown that if the Pluckers's co-ordinates of S be calculated with the help of a different set of basis vectors of S, then these may differ from the above Plucker's co ordinates by a constant multiplying factor. It can also be shown that under a change of basis of the linear space V, the Plucker's co-ordinates of the linear subspace S either remain the same, or are multiplied by a constant. Hence the Plucker's co-ordinates of a linear subspace S of V are determined uniquely, except for an arbitrary mutiplicative (non-null) constant factor. Thus the Plucker's co-ordinates are *homogeneous* co-ordinates.

Exercises

1. Show that, by a change of basis of the underlying linear space V of rank n, the components of any skew-symmetric tensor with n indices are multiplied by the determinant of the matrix associated with the change of basis.

2. Show that the only skew-symmetric multilinear function in n vectors $\xi, \eta, \dots, \zeta$, on a linear space V of rank n, is the determinant formed by the co-ordinates of these vectors, referred to some basis of V, or any constant multiple of the determinant.

3. Verify that the components of the skew-symmetric tensor G, as defined above, are given by the kth order minors of the matrix (1), when referred to the particular basis, with respect to which the co-ordinates of the vectors are taken.

Show that the components of G are multiplied by a constant factor, by a change of the basis of the underlying linear space V.

4. Show that the Plucker's co-ordinates of a linear subspace S, calculated with the help of two different bases of S, differ at most by a constant multiplying factor.

BIBLIOGRAPHY

Aitken, A.C., *Determinants and Matrices*, Oliver Boyd, 1962.

Albert, A.A., *Modern Higher Algebra*. *Introduction to Algebraic Theories*, The University Press of Chicago, 1937, 1941.

Amir-Moez. A. R. and Fass A. L., *Elements of Linear Spaces*, Pergamon Press, 1962.

Birkhoff, G. and MacLane, S., *A Survey of Modern Algebra*, New York, Macmillan, 1965.

Bocher, M., *Introduction to Higher Algebra*, New York, Macmillan, 1907.

Burnside, W.S. and Panton, A. W., *The Theory of Equations*, Vol.II, Dover Publ. Inc., 1960.

Dickson, L.E., *Algebraic Theories*, Dover Publ. Inc., 1959.

Finkbeiner(II), D. T., *Introduction to Matrices and Linear Transformations*, San Fransisco, W. H. Freeman & Co., 1960.

Gantmacher, F.R., *The Theory of Matrices*, Vol. I, (Eng. Trans. by Hirsch, K.A.) Chelsea Publ Inc., 1960 ; *Applications of the Theory of Matrices* (Eng. Trans by Brenner, J. L.) Interscience Publ. Inc., 1959.

Gel'fand, I.M., *Lectures on Linear Algebra* (Eng. Trans. by Shenitzer, A.) Interscience Publ. Inc., 1961.

Greub, W.H , *Linear Algebra*, Berlin, Springer-Verlag, 1963.

Halmos, P. R., *Finite Dimensional Vector Spaces*, D. Van Nostrand Co. Inc., 1958

Hamburger, H.L. and Grimshaw, M.E., *Linear Tranformations*, New York, Cambridge University Press, 1951.

Hoffman, K. and Kunze, R., *Linear Algebra*, Prentice Hall, 1961.

Hohn, F.E., *Elementary Matrix Algebra*, New York, Macmillan, 1964.

Jacobson, N., *Lectures on Abstract Algebra*, Vol. II, D. Van Nostrand Co. Inc., 1963.

Kelley, J. L., *Introduction to Modern Algebra*, D. Van Nostrand Co. Inc., 1960.

Lang, S., *Algebra*, *Linear Algebra*, Addison-Wesley Publ. Co. Inc., 1965, 1966.

Levi, F. W., *Algebra*, Vol I, The Calcutta University, 1942.

Lewis, D. J., *Introduction to Algebra*, New York, Harper & Row, 1965.

MacDuffee, C. C., *An Introduction to Abstract Algebra*, John Wiley & Sons Inc., 1940 ; *Vectors and Matrices* (*Carus Mathematical Monograph* No. 7), The Mathematical Association of America, 1943 ; *The Theory of Matrices*. Chelsea Publ. Co., 1946.

Mal'cev, A. I., *Foundation of Linear Algebra* (Eng. Trans by Brown. T.C.) San Fransisco, W.H. Freeman & Co., 1963.

MacCoy, N.H., *An Introduction to Modern Algebra*, Bostan, Allyn and Bacon Inc., 1965.

Mirsky, L *An Introduction to Linear Algebra*, New York, Oxford University Press, 1955.

Mostow, G. D., Sampson, J. H. and Meyer, J., *Fundamental Structures of Algebra*, MacGraw-Hill Book Co. Inc., 1963.

Perlis, S., *Theory of Matrices*, Addison-Wesley Publ. Co. Inc., 1952.

Racine, C., *Introduction to Abstract Algebra*, Madras, S Viswanathan, 1957.

Schreier, O. and Sperner, E., *Modern Algebra and Matrix Theory* (Eng. Trans. by Davis, M and Hausner, M.), Chelsea Publ. Co., 1961.

Shanti Narayan, *A Text Book of Matrices*, New Lelhi, S. Chand & Co., 1964.

Shields, P. C., *Linear Algebra*, Addison-Wesley Publ. Co. Inc., 1965.

hilov. G. E., *An Introduction to the Theory of Linear Spaces* (Eng. Trans. by Silverman, R. A.) Prentice Hall, 1961.

mirnov, V. I., *Linear Algebra* (*A Course of Higher Mathematics*, Vol. III-1)(Eng Trans. by Brown, D.E), Pergamon Press, 1964

tewart, F.M, *Introduction to Linear Algebra*, D. Van Nostrand Co. Inc., 1963.

toll., R R., *Linear Algebra and Matrix Theory*, MacGraw-Hill Book Co. Inc., 1952.

hrall, M and Tornheim, L., *Vector Spaces and Matrices*, John Wiley, & Sons Inc., 1957.

urnbull, H. W. and Aitken, A. C, *An Introduction to the Theory of Canonical Matrices*, Dover Publ. Inc., 1961.

Vaerden, B. L. van der, *Modern Algebra*, Vol. II (Eng. Trans. by Benac, T. J.) New York, Frederick Ungar Publ. Co., 1950; *Gruppen von Linearen Transformation*, Chelsea Publ. Co., 1948.

edderburn, J H M, *Lectures on Matrices*, Dover Publ. Inc., 1964.

efimov, N.V., *Quadratic Forms and Matrices* (Eng. Trans. by Shenitzer, A) Academic Press, 1964.

urmuhl, R., *Matrizen and ihre technischen Anwendungen*, Berlin, Springer-Verlag, 1961.

———

INDEX

Adjoint of a matrix, 99
adjoint operator, 249
adjugate of a matrix, 99
alternation, 298
alternatization, 298
automorphism, 240

Basis, 33
Bessel's inequality, 52
bijective mapping, 5
Binet-Cauchy theorem, 104
bi-orthogonal bases, 280
block multiplication, 59

Cartesian product of sets, 2
Cauchy-Schwarz inequality, 44, 104
Cayley's formulae, 261
Cayley-Hamilton theorem, 156
characteristic :
  polynomial, 156, 236
  root, 161
  subspace, 166
  value, 161
  vector, 165
characteristic of a ring, 13
cofactor, 97
companion matrix, 157
complement, 2, 35
composition, 6
congruence of :
  matrices, 132
  quadratic forms, 136
  skew symmetric matrices, 148
  symmetric bilinear forms, 147
  symmetric matrices, 132
co-ordinates, 36, 51
  multilinear function, 288
  Plucker, 299
  vector, 38
conjugate space, 275
co-ordinate transformation :
  covariant vector, 282
  contravariant vector, 282
  dual basis, 282
  linear function, 277
  linear operator, 234
  multilinear function, 288
  vectors, 38, 50
Crammer's rule, 119

Decomposition, polar, 260
  spectral, 269
determinant, 88
determinant divisor, 126
dimension, 34
dimensionality, 31
direct sum, 35,
disjoint sets, 1
divisibility, 18
division algorithm, 20,153
divisors of zero, 12
domain :
  Euclidean, 20
  Gaussian, 19
  integral, 12
  Noetherian, 19
  uniqely factorisable, 19
dual basis, 281
  space, 275

Eigen value, 161, 236
  vector, 165, 236
element, 1
  algebraic, 24
  irreducible, 19
  null, 11
  prime, 19
elementary divisor, 128
equivalence :
  of bilinear forms, 122
  of matrices, 62, 65, 121, 123
  column-, 65
  row-, 62
equivalence relation, 3
extension of a field, 13

  algebraic, 24
  simple, 24
  quadratic, 24

Factorisation, 19
field, 12
  algebraically closed, 24
  formally real, 25
  Galois 16
  quotient, 14
  skew-, 12
  splitting, 23
  sub-, 13
forms :
  bilinear, 122
  definite, 142
  Hermitian, 151, 213
  indefinite, 142
  linear, 275
  quadratic, 136, 216, 229
  symmetric bilinear, 147, 215, 229

function, 5
  bilinear, 236, 296
  linear, 275
  mixed bilinear, 286
  multilinear, 286
fundamental theorem :
  of classical algebra, 24
  of general algebra, 23

Generators, 28
Gramm-Schmidt orthogonalisation process, 49
group, 8
  abelian 8
  commutative, 8
  factor-, 11
group of matrices :
  full linear, 77
  general linear, GL(n,K), 77
  orthogonal, 80
  unimodular, 104
  unitary, 85

Hadamard determinant theorem, 104
Hermitian congruence, 150
  forms, 151, 213
  matrix, 84, 150
  operator, 251
highest common factor, 20
homomorphism, 17
  kernel of, 17
homomorphism theorem :
  for groups, 17
  for rings, 17

Ideals, 16
injective mapping, 5
intersection of sets, 1
intersection of planes in Euclidean space, 116
invariant, factor, 127
  subspace 240
  vector, 165
isomorphism, 17

Jacobi's theorm, 102
J/(g), 4, 15
Jordan :
  block, 160
  generalised normal form, 203
  matrix, 196
  normal form, 203

Lagrange's method of reduction, 139

Lagrange's theorem, 11
Laplace's development, 100
linear :
  dependence, 29
  forms, 275
  function, 275
  functional, 275
  independence, 29
  space, 26
  ......, Euclidean, 53, 255
  ....., normed, 42
  ....., unitary, 43
  subspace, 26
  sum, 27
linear equations :
  homogeneous system, 109
  non-homogeneous system, 112
linear operator, 231

- adjoint of, 249
- automorphism, 240
- characteristic polynomial, 236
- eigen value, 236
- eigen vector, 236
- Hermitian, 251
- ......, non-negative, 253
- invariant subspace, 240
- invertible, 231
- kernel, 238
- matrix representation, 233
- minimal polynomial, 236
- non-singular, 239
- normal, 251
- nullity, 238
- null-space, 238
- orthogonal, 264
- projection, 266
- range, 237
- rank, 237
- self-adjoint, 251
- skew-Hermitian, 251
- skew-symmetric. 255
- symmetric, 255
- unitary, 258

Mappings, 5
matrix, 55
  - block diagonal, 191
  - characteristic, 156
  - column-space, 67
  - companion, 157
  - complex, 84, 168
  - congruent, 132
  - decomposable, 192
  - derogatory, 190
  - diagonable, 187
  - diagonal 73
  - echelon, 62
  - elementary, 74
  - equivalent, 121, 123
  - Hermitian, 84. 209
  - idempotent, 83
  - identity, 73
  - indecomposable, 192
  - monomial, 82
  - nilpotent, 82
  - non-derogatory, 160
  - non-singular, 75
  - normal, 86
  - null, 73
  - nullity of, 71
  - operations on, 55-59
  - orthogonal, 80,216
  - orthogonally congruent, 215
  - orthogonally similar, 215
  - permutation, 81
  - rank of, 68
  - real, 176
  - row-space, 66
  - scalar, 73
  - semi-decomposable, 243
  - similar, 183
  - singular, 76
  - skew-Hermitian, 84
  - skew-symmetric, 79
  - square, 72
  - symmetric, 79, 215, 228
  - triangular, 82, 195
  - unitarily congruent, 208
  - unitarily similar, 208
  - unitary, 85,209

minor, 97, 106
multiplicity of an eigen value :
  - algebraic, 167
  - geometric, 167

Neumann theorem, 45
norm of a vector, 42
normal form of matrices :
  - classical canonical, 196
  - fully reduced, 65
  - fully swept out, 64
  - generalised Jordan, 203
  - Jacobson canonical, 203
  - Jordan, 196
  - natural, 201
  - quasi-natural, 202
  - rational canonical, 201
  - Smith's, 124

normed linear space 42
nullity :
  - of a matrix, 71
  - of an operator, 238

Operator, 6
  - linear, 231

orthogonality, 47
orthogonal complement, 49
orthogonal similarity, 215-219
  orthogonal matrices, 216
  quadratic forms, 216
  symmetric bilinear forms, 216
  symmetric matrices, 215
orthonormal basis, 48

Partition of a set, 3
permutation, 6
Perseval's identity, 52
Plucker co-ordinates, 299
polynomial, 20
  characteristic, 156, 236
  matric, 153
  matrix as a root of, 154
  minimal, 158, 236
  non-separable, 23
  roots of, 23
  separable, 23
product :
  inner, 43
  matrix, 57
  operator, 234
  scalar, 43
  tensor, 292
proper values, 161
  vectors, 165

Quadratic form, 136
  index, 138
  signature, 138
quadrics :
  affine classification, 224
  Euclidean classification, 220
  projective classification, 144
quasi-metric space, 53
quasi-normed space, 52

Rank of a linear space, 34
  of a matrix, 68, 106
  of a tensor, 291
replacement theorem, 29
residue classes, 4, 15, 17
Riesz representation theorem, 278

ring, 12
  commutative, 12
  Noetherian, 16
  polynomial, 20
  principal ideal, 16
  quotient, 16
  sub-, 13

Scalar, 26
Segre characteristic, 198
set of residue classes, 4
similarity of matrices, 183
  to a diagonal matrix, 186
  orthogonal-, 215
  unitary-, 208
skew field, 12
skew-Hermitian, matrix, 84, 151
  operators, 251
Smith's normal form, 124
space :
  linear, 26
  pseudo-normed, 53
  quasi-normed, 32
  unitary, 43
  vector, 26
spectral decomposition, 269
spectral theorem, 272
subgroup, 8
  cosets of, 10
  cyclic, 10
  normal, 11
surjective mapping, 5
Sylvester's law of inertia, 139
symmetric matrix (real) :
  congruence, 134
  index, 134
  signature, 134
symmetrization, 297
system of double composition, 11
  of single composition, 7

Tensor,
  alternation, 298
  anti-symmetric, 297
  components, 291
  contraction, 292
  contravariant, 290
  covariant, 290
  lowering of indices, 294
  metric, 296

    rank of, 291
    raising of indices, 295
    skew-symmetric, 297
    symmetric, 297
    symmetrization, 297
    valence of, 291
total matric algebra, 72
transformation, 6
    non-singular, 6
    singular, 6
transpose of a matrix, 58
type (*p*, *q*), 286

Union of sets, 1
unit, 18
unitary space, 43
unitary similarity, 208-214
    Hermitian forms, 213
    Hermitian matrices, 209
    normal matrices, 210
    skew-Hermitian matriees, 212
    unitary matrices, 207

Valence of :
    a multilinear function, 286
    a tensor, 291
vector, 26
    characteristic, 165
    contravariant, 280
    covariant, 280
    eigen, 165, 236
    invariant, 165
    length of, 44
    norm of, 44
    normalised, 48
    null, 26
    orthogonal set, 48
    orthogonality, 47
    orthogonal prejection, 50
    orthonormal system, 48
    proper, 165
    unit, 48
vector space, 26
    sub-, 26

THE END